IBM HR Analytics Employee Attrition & Performance

Data Evangelists

Omkar Bapat | Vishal Budhani | Mayuri Deshpande | Sachin Singh | Shipra Shinde

ITCS 6162 Knowledge Discovery in Databases

University of North Carolina Charlotte

Date: 06/28/2017

Abstract

The project focuses on analyzing the data to unravel the factors that lead to employee attrition within IBM. We followed CRISP - DM process to analyze and relate various attributes leading to employee attrition. We used various charts ( bar charts, histograms etc.) to visualize and understand the data properly during Exploratory Data Analysis (EDA) phase. Data understanding helped us in later phases wherein we were able to classify the data more efficiently and hence were able to generate more accurate predictions and models for the study.

We have concluded that “DistanceFromHome, OverTime, TrainingTimesLastYear & WorkLifeBalance” are the prime reasons for high attrition rate. We have classified these attributes into High Attrition Cluster and have given various statistical data supporting our final model.

This research work can be taken forward and certain business decisions can be taken around the responsible factors for reducing the attrition rate.

IBM HR Analytics Employee Attrition & Performance

We are presenting this paper as part of our final project work for Knowledge Discovery of Databases coursework at University of North Carolina at Charlotte.

HR Analytics is the application of data mining and business analytics techniques to Human Resources data. The goal of the analytics is to provide an organization with insights for managing the employees effectively so that the organizational goals are met efficiently. We will be exploring IBM HR Analytics employee data to predict some of the major factors which influence employee's decision making about staying or leaving the organization.

We will follow the Cross-Industry Standard Process for Data Mining (CRISP-DM) to build predictive models using predictor(s) available in the data set and present it as findings for the attrition rate. The models can give insights to Human Resources department about the probable causes for the attrition within the company.

# **Dataset**

**Description and Purpose**

The key to success of any organization is attracting and retaining top talents. This is one of the major responsibilities of the Human Resources department within an organization. One of the most important and challenging task for any Human Resources department is to determine factors which influence employee's decision making about staying or leaving an organization. Therefore, it is of vital importance to visualize factors pivotal in employee exit. The dataset showcases information about IBM employee attrition considering the basic employee information as well as employee work environment analysis. The dataset consists of 35 fields containing employee information like the job satisfaction rating, percent salary hike, performance rating, years at company, etc. which is important in determining the employee attrition and performance.

**Source**

[Kaggle - https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset.](https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset) **Related Work**

Document any related work – include references to the work if applicable.

# **Business Research/Understanding**

**Project Objectives**

**Problem Domain**

Human Resources (HR) analytics involves the application of analytic processes within a Human Resources department for improving employee performance. HR analytics enable organizations to use their wealth of employee data to make better decisions about their workforces and improve operational performance. The problem domain encompasses analysis of employee attrition so as to figure out reasons for people leaving/staying in the company.

**Requirements**

IBM HR Analytics Employee Attrition & Performance dataset from Kaggle along with following tools: R, RStudio, Microsoft-Excel.

**Restrictions**

Time

**Data Mining Problem Definition**

The goal of our research is to visualize factors that leads to employee attrition. The analysis consists of a full EDA and predictive analysis of attrition at IBM, followed by some calculations of the likely business costs under each of the various models we will use. Any modern-day firm always look for cutting down various costs (Onboarding, hiring costs etc.) that they incur on hiring new employees & hence various HR departments look for various analytical ways to model out the employee attrition and figure out the most prominent factors responsible for it. In our study, a highly accurate model doesn’t automatically guarantee a higher Return on Investment (ROI), however it provides a deep insight to HR department to analyze the predicted models and then to take necessary steps based on the models generated.

**Strategy**

We plan to proceed in the following fashion

1. We will understand the underlying business requirements in detail and will set

objectives for our research work.

2. By using Exploratory Data Analysis we will familiarize ourselves with the dataset

available and will discover initial insights.

3. Cases and variables appropriate for the analysis will be selected and analyzed using

various R operations available.

4. We will apply multiple techniques to model the data correctly.

5. We will evaluate the model(s) generated against our business requirements and

objectives.

6. We will be publishing the model(s) and inferences using data handling and

visualization tool like Tableau.

# **Data Understanding**

**Exploratory Data Analysis**

**Description of the Data.**

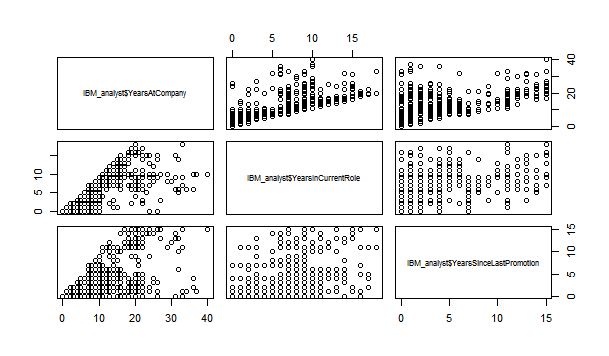
|  |  |  |  |
| --- | --- | --- | --- |
| **Serial Number** | **Variable Name** | **Variable Type** | **Comments** |
| 1 | Age | Continuous | Age shows the employee age in years |
| 2 | Attrition | Categorical | Target. Indicator of whether the customer has left the company (Yes  or No) |
| 3 | BusinessTravel | Categorical | Business Travel shows the travelling tendency of the employee |
| 4 | DailyRate | Continuous | The rate of pay for a day's work by an employee |
| 5 | Department | Categorical | Department to which an employee belongs to ( e.g. Human Resources, Sales etc.) |
| 6 | DistanceFromHome | Continuous | Distance of office from home |
| 7 | Education | Continuous | Educational level/qualification of employee on a scale of 1-4 |
| 8 | EducationField | Categorical | Education field to which the employee belongs to. |
| 9 | EmployeeCount | Continuous | Unary Field |
| 10 | EmployeeNumber | Continuous | Unique ID of the employee |
| 11 | EnvironmentSatisfaction | Categorical | Work Environment happiness factor on a scale of 1-4 |
| 12 | Gender | Categorical | Gender of an employee (Male/Female) |
| 13 | HourlyRate | Continuous | Hourly charges of an employee |
| 14 | JobInvolvement | Discrete | Allocates a score on scale 1-4 on the basis of how effective is the employee contribution |
| 15 | JobLevel | Discrete | Designation hierarchy field on scale of 1-4 |
| 16 | JobRole | Categorical | Designated job position of the employee |
| 17 | JobSatisfaction | Discrete | Allocates a score on scale 1-4 on the basis of how satisfied the employee is with the job |
| 18 | MaritalStatus | Categorical | Married status of the employee (Single, Married, Divorced) |
| 19 | MonthlyIncome | Continuous | Specifies the monthly income |
| 20 | MonthlyRate | Continuous | The rate of pay for a month's work by an employee |
| 21 | NumCompaniesWorked | Continuous | Specifies the number of companies that the employee has worked in |
| 22 | Over18 | Categorical | Unary |
| 23 | OverTime | Categorical | States whether the employee worked overtime or not. Takes values as yes/no |
| 24 | PercentSalaryHike | Continuous | salary hike in percentage |
| 25 | PerformanceRating | Discrete | Performance evaluation.Takes a score of 3 or 4 |
| 26 | RelationshipSatisfaction | Discrete | Allocates a score on scale 1-4 on the basis of how satisfied the employee is with his colleagues |
| 27 | StandardHours | Discrete | Standard working hours for employees. Unary field having value 80 |
| 28 | StockOptionLevel | Discrete | The level of stock option plan assigned to the employee ranging from 0 to 3 |
| 29 | TotalWorkingYears | Continuous | The total number of years that the employee is working |
| 30 | TrainingTimesLastYear | Continuous | The number of trainings that the employee has completed in the last year |
| 31 | WorkLifeBalance | Discrete | The work life balance rating given by the employee ranging from 1to 4 |
| 32 | YearsAtCompany | Continuous | Number of years the employee has been working at the company |
| 33 | YearsInCurrentRole | Continuous | Number of years the employee is in the current role |
| 34 | YearsSinceLastPromotion | Continuous | Number of years since the last promotion of the employee |
| 35 | YearsWithCurrManager | Continuous | Number of years that the employee has been working with the current manager |

**Estimation:**

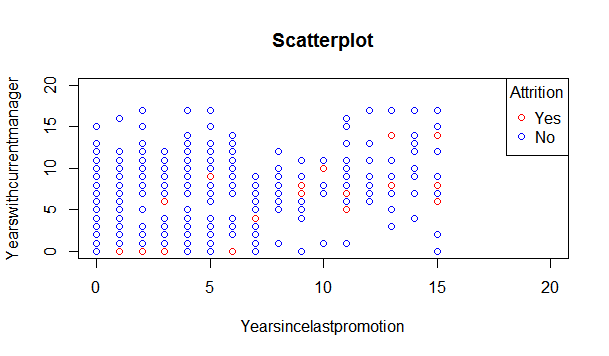
As a part of the exploratory data analysis we categorized our data and identified certain patterns and trends through observations and center of measure methods.We used barchart for categorical and discrete values and histograms scatter plot for numeric variables.Through the summary command in R we analysed the dataset further thereby helping the graphical analysis.We have also done regression analysis to study and understand how various factors of the dataset are co -related and how we can draw inferences.

Exploratory data analysis helped us with understanding of interesting relations within the dataset which we have stored in the form of subsets.These subsets will help us understand the reason of attrition.

Histogram for all the variables are available in Appendix A.



Scatterplot matrix for three important factors as 1.years at company 2.years in current role 3. years since last promotion



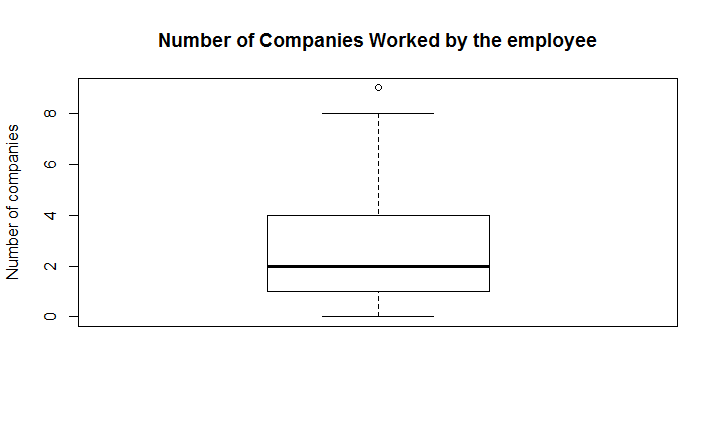
Scatterplot matrix for year since last promotion and years with manager in relation with attrition

**Summaries**

Summaries of all the variables is available in appendix B.

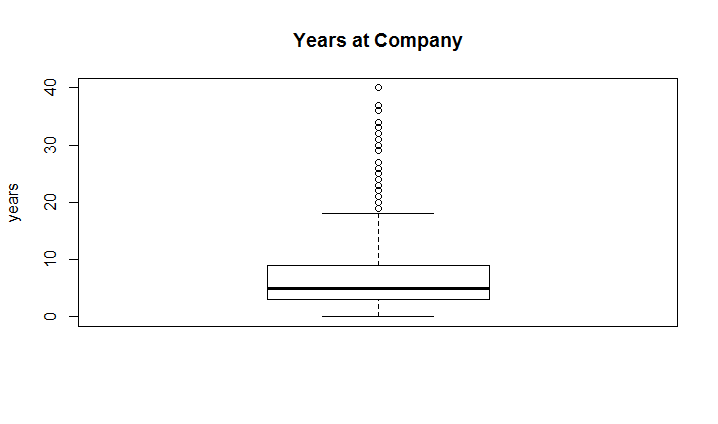
**Identification of Outliers:**

**Variable Name: NumCompaniesWorked**



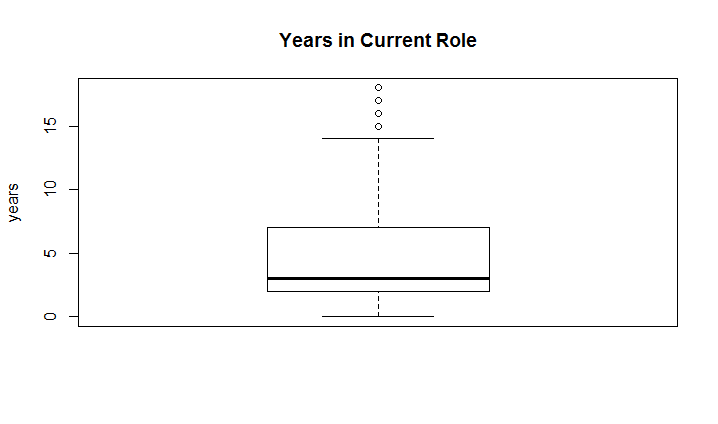
**Outliers Values:** 9

**Variable Name: YearsAtCompany:**



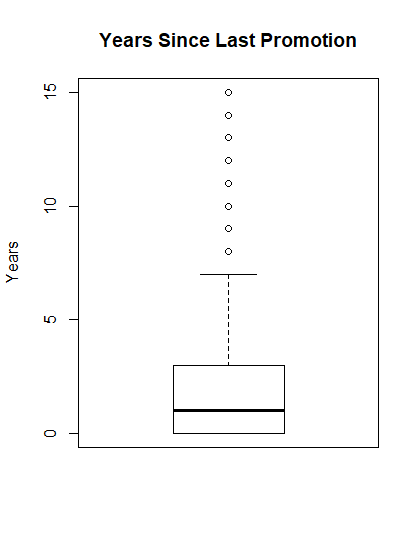
**Outlier values:** 19, 20, 21, 22, 23, 24, 25, 26, 27, 29, 30, 31, 32, 33, 34, 36, 37, 40 years

**Variable Name: YearsInCurrentRole:**



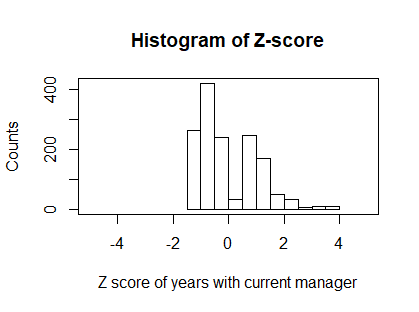
**Outlier Values:** 15,16, 17, 18 years

**Variable Name: YearsSinceLastPromotion:**



Outlier values: 8, 9, 10, 11, 12, 13, 14, 15 years

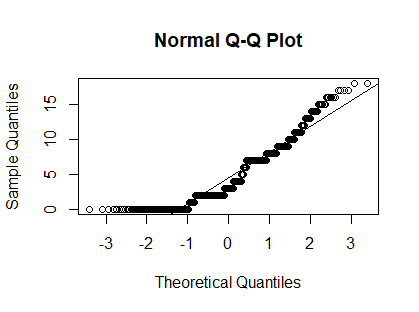
**Variable Name: YearsWithCurrManager:**



Outlier values: 15, 16, 17 years

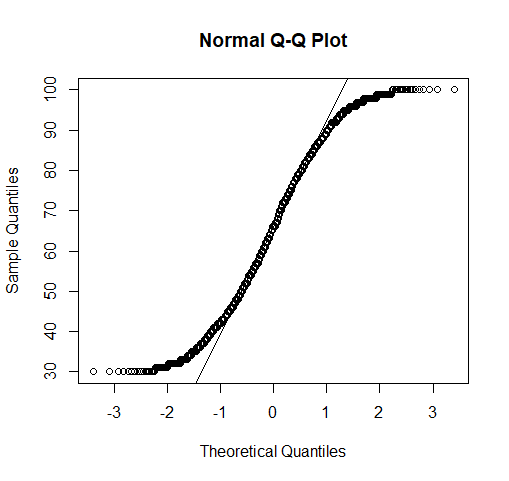
**Q-Q Plots**

**Years In Current Role**



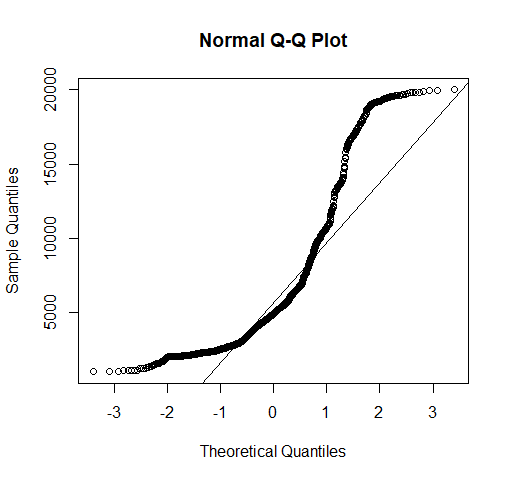
The distribution for the years in current role is not normal.

**Hourly Rate**



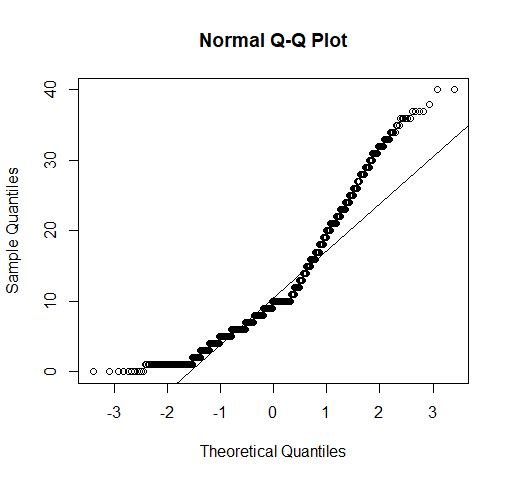
The distribution is not normal

**Monthly Income**



The distribution is not normal.

**Total Working Years**



The distribution is not normal.

**Data Quality.**

We have used EDA for understanding the dataset and for finding the relationship between various variables, finding interesting subsets, etc. With the help of primary observations from the above diagrams and the listed summaries of the variables we have found a outliers in the variables of the dataset. We have also found unary variables which we will remove in the data preparation phase.

One of the most important part is the missing data but the dataset does not have any missing data. We can see that most of the variables do not show normal distribution.

**Regression Analysis:**

1. **Regression Analysis for Monthly Income :** We have done a regression analysis of monthly income against TotalWorkingYears, NumCompaniesWorked & YearsAtCompany.

Simple regression line equation for it is:

Monthly Income = 1277.93 + 457.67(TotalWorkingYears) -48.92(NumCompaniesWorked)

+27.74(YearsAtCompany).

Regression analysis data is as follows:

Call:

lm(formula = monthly\_income ~ total\_workingyears + number\_companiesworked +

years\_atCompany)

Residuals:

Min 1Q Median 3Q Max

-10922.5 -1744.3 -50.1 1375.0 11360.9

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1277.93 156.44 8.169 6.64e-16 \*\*\*

total\_workingyears 457.67 14.05 32.567 < 2e-16 \*\*\*

number\_companiesworked -48.92 34.30 -1.426 0.154

years\_atCompany 27.74 17.46 1.589 0.112

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2983 on 1466 degrees of freedom

Multiple R-squared: 0.5993, Adjusted R-squared: 0.5985

F-statistic: 730.8 on 3 and 1466 DF, p-value: < 2.2e-16

We can see that the adjusted RSquare value is approximately 60%. In order to make it better, we will add few more predictors to it and will try to improve the value of MonthlyIncome for our regression model.

**Subsets.**

Subsets where employees tend to leave the company (237 employees have left the company):

1. Employees who are promoted within the last 5 (200 employees out of 237)
2. Employees of the age ranging from 25 to 35 (122 employees)
3. Employees working with the current manager for less than or equal to 2 years (146 employees)
4. Employees working with the current manager for less than or equal to 3 years and promoted within last 2 years and age less than 30 (82 employees)

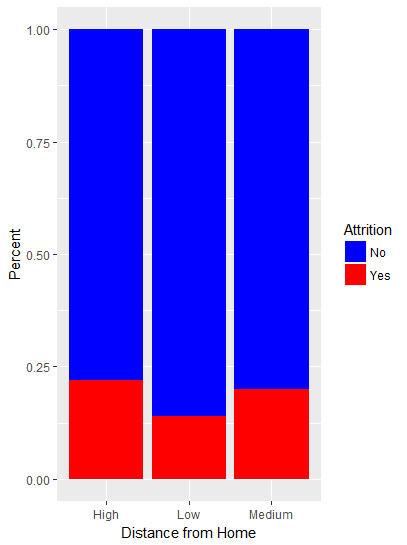
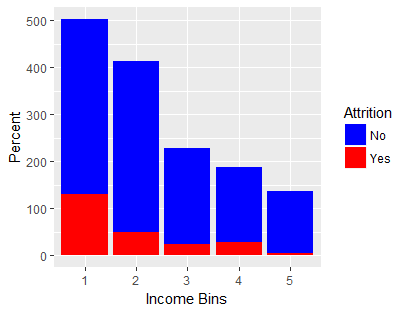
# **Data Preparation**

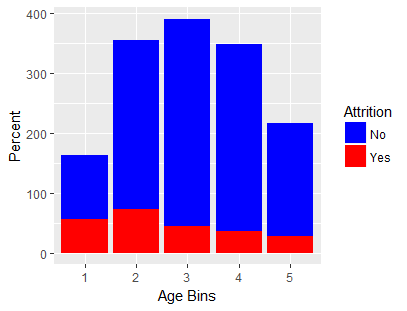
**Cleansing.** (removal of anomalous attributes, features not useful to research, etc.)

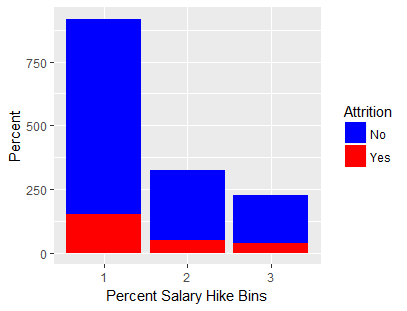
As part of data cleansing, we will be removing the following attributes. The reasons for removing the attributes are mentioned below as well.

1. EmployeeCount: Employee Count field can be removed as it contains only one value that is “1”. hence for our data preparation we can remove this attribute.
2. Over18: Over18 attribute can be removed as it contains only one value that is “Y”. This field checks if the employee is above 18 years of age. This seems to be quite obvious that every employee will be above 18 years of age hence this attribute does not help us in prediction, therefore it’s best suited to remove it.
3. EmployeeNumber: EmployeeNumber is an ID field hence will not help much in prediction, therefore we can remove it.
4. StandardHours: StandardHours is a unary variable having value “80”. The standard working hours for an employee is 80 hours. So, the variable can be removed from the dataset.

**Binning and Discretization.**

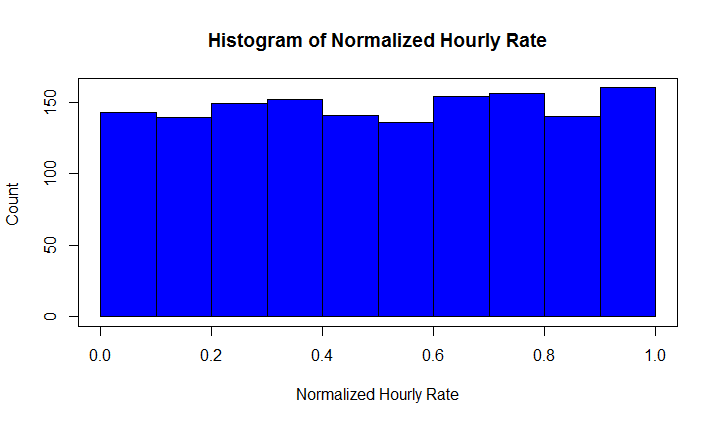
1. DistanceFromHome: The variable ‘DistanceFromHome’ can be binned into 3 different bins.  
   Values ranging from 1 to 10 - Low  
   Values ranging from 11 to 20 - Medium   
   Values ranging from 21 to 30 - High  
   
2. MonthlyIncome: Discretizing monthly income into 5 categories using k-means clustering.  
   Bin: 1 2 3 4 5  
   Levels: 1009 3722 6062 9355 14732 19999  
   
3. Age: Discretizing age into 5 categories using k-means clustering.  
   Bin: 1 2 3 4 5  
   Levels: 18 26 32 38 47 60

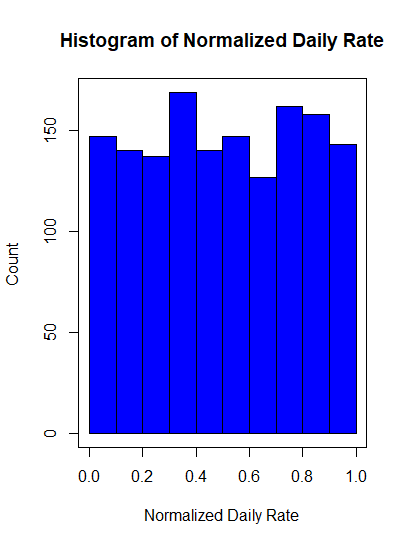


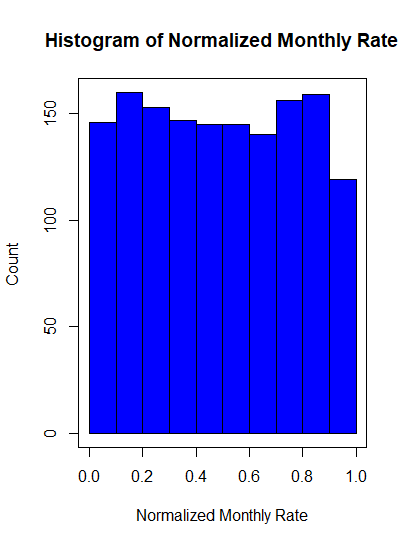
1. PercentSalaryHike: Discretizing percent salary hike into 3 categories using k-means clustering.  
   Bin: 1 2 3  
   Levels: 11 16 20 25  
   

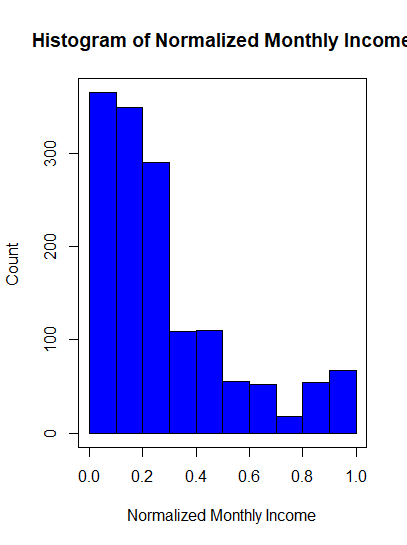
**Missing Values.** IBM HR Analytics Dataset does not have any missing values.

**Normalization of Numeric Variables.** We have normalized HourlyRate, DailyRate, MonthlyRate, MonthlyIncome using min max normalization.



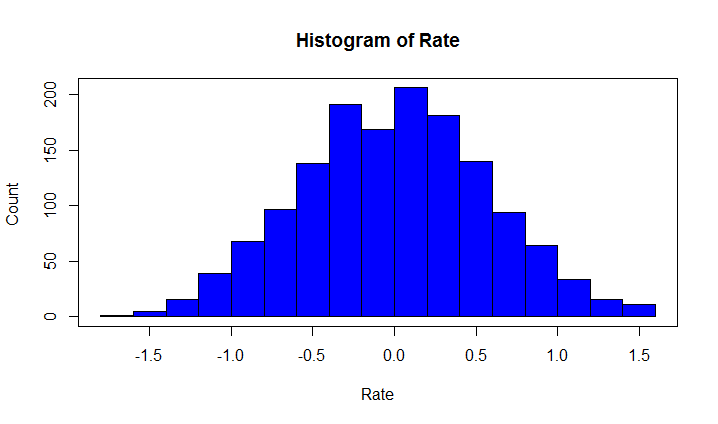






**New Variables.**

Combining HourlyRate, DailyRate, MonthlyRate into a single variable “Rate”.



**Other transformations.**

Converting the dichotomous categorical variable ‘Attrition’ into flag variable using following conversion: ‘Yes’ as 1 and ‘No’ as 0.

# **Modeling**

We have used Classification and Regression Trees (CART), C5.0, k-Nearest Neighbor Algorithm, Random Forest, Neural Network to find the best possible prediction model. For classification, we used CART, Random Forest which also gives the rank of important variables. Apriori algorithm is used to find the association rules.

**Prediction**

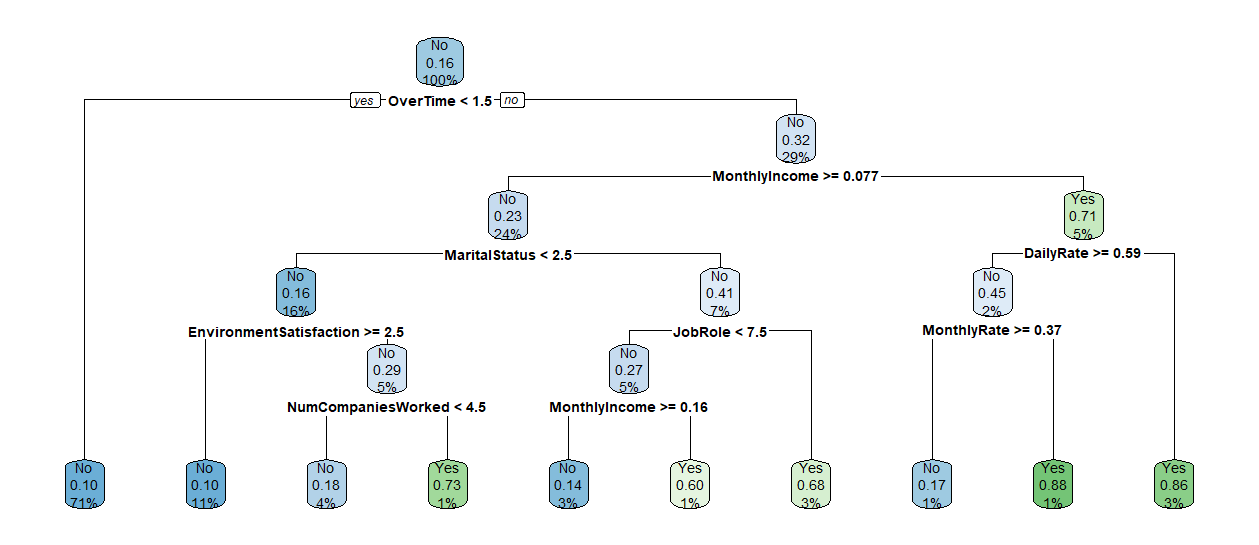
The dataset is partitioned into training and test dataset using two fold cross validation.

All the numeric variables are normalized using min-max normalization, categorical variables are converted into numeric and used for prediction modeling.

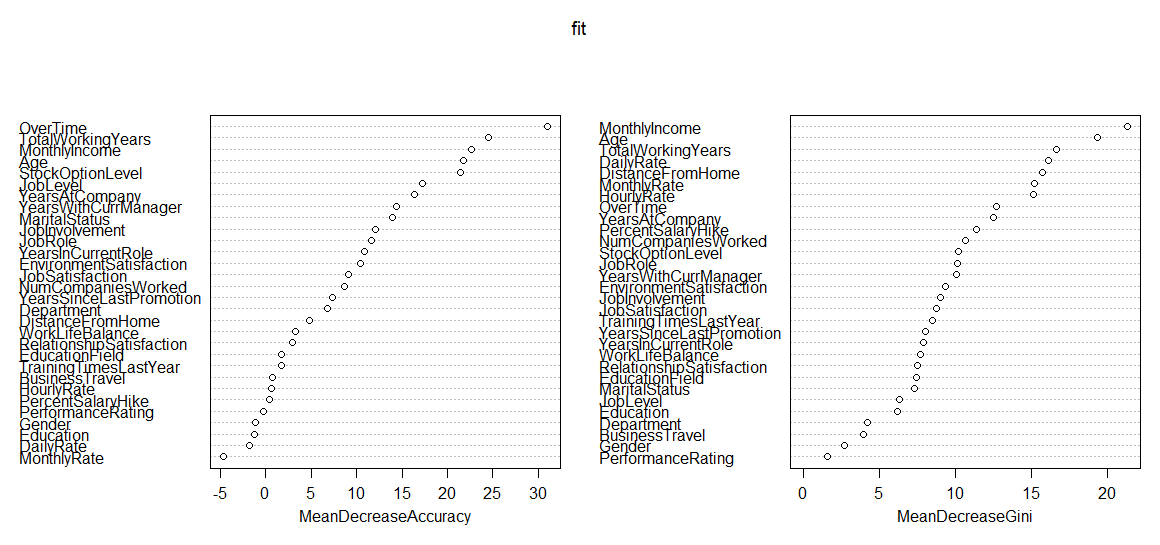
The models are trained using CART, C5.0, k-Nearest Neighbor, Random Forest, Neural Network algorithms. Models are tested by predicting the results for the test data set.

**Classification**

Classification and Regression Trees (CART)



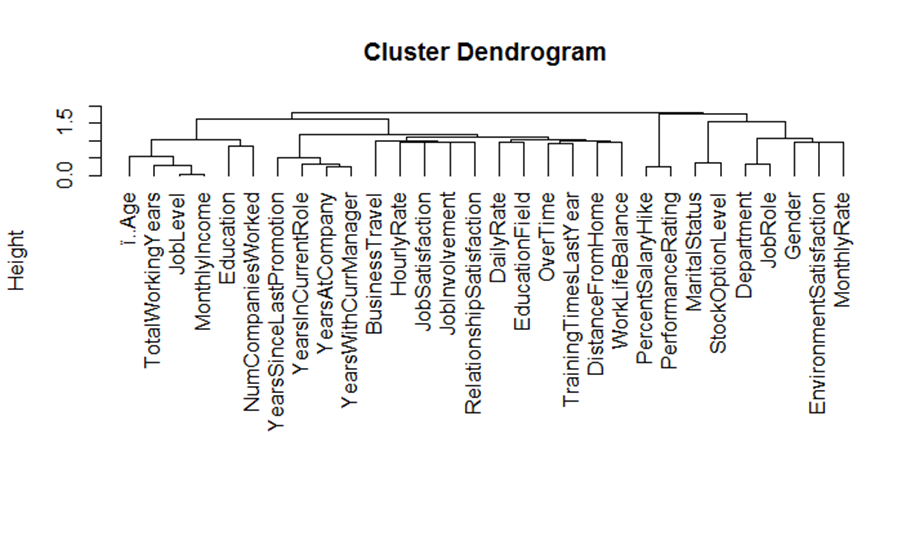
Random Forest



The output of both the algorithms show the most important factors affecting the attrition namely OverTime, MonthlyIncome, Age, TotalWorkingYears, MaritalStatus.

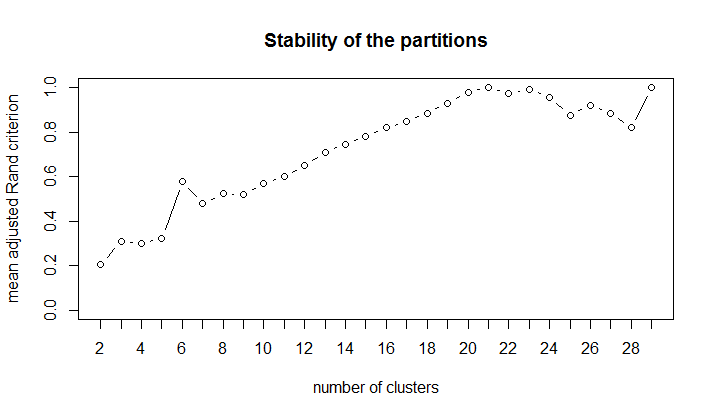
**Clustering**

We performed variable clustering on the dataset available for the study. We performed transformation from categorical to numerical variables of BusinessTravel, Department, Gender, MaritalStatus, OverTime, JobRole, EducationField and excluded the target variable Attrition in order to achieve variable clustering.



**Tree of Numerical Variable groups.**

The dendrogram suggests that the 35 input variables can be combined into approximately 9 – 11 groups of variables. For determining the number of clusters we plot an index of stability of partitions of variables.



The plot of stability of variable cluster partitions suggests approximately a 7 to 9-cluster solution.

The list of all 9 clusters along with the variable names are given below:

|  |  |  |
| --- | --- | --- |
| **Profile Description Cluster**  Age  Education  JobLevel  MonthlyIncome  NumCompaniesWorked  TotalWorkingYears | **Job Indicators Cluster**  BusinessTravel  HourlyRate  JobInvolvement  JobSatisfaction  RelationshipSatisfaction | **Daily Rate & Education Cluster**  DailyRate  EducationField |
| **Office Centric Cluster**  Department  JobRole | **High Attrition Cluster**  DistanceFromHome  OverTime  TrainingTimesLastYear  WorkLifeBalance | **Personal Attributes Cluster**  EnvironmentSatisfaction  Gender  MonthlyRate |
| **Stock & Marital Status Cluster**  MaritalStatus  StockOptionLevel | **Performance Cluster**  PercentSalaryHike  PerformanceRating | **Duration Indicator Cluster**  YearsAtCompany  YearsInCurrentRole  YearsSinceLastPromotion  YearsWithCurrManager |

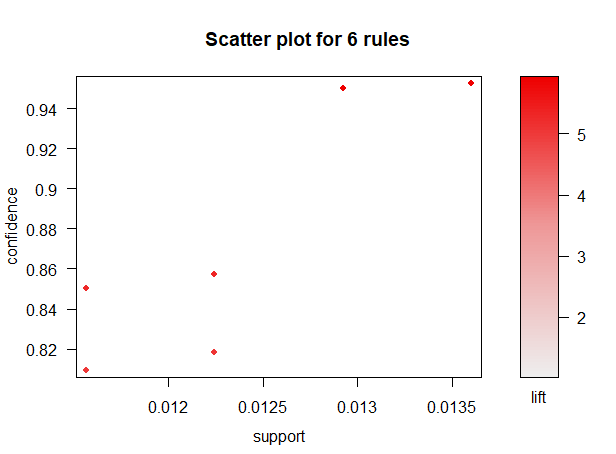
Analysing the clusters, **High Attrition Cluster** having “DistanceFromHome, OverTime, TrainingTimesLastYear, WorkLifeBalance” in it plays an important role in Employee Attrition.

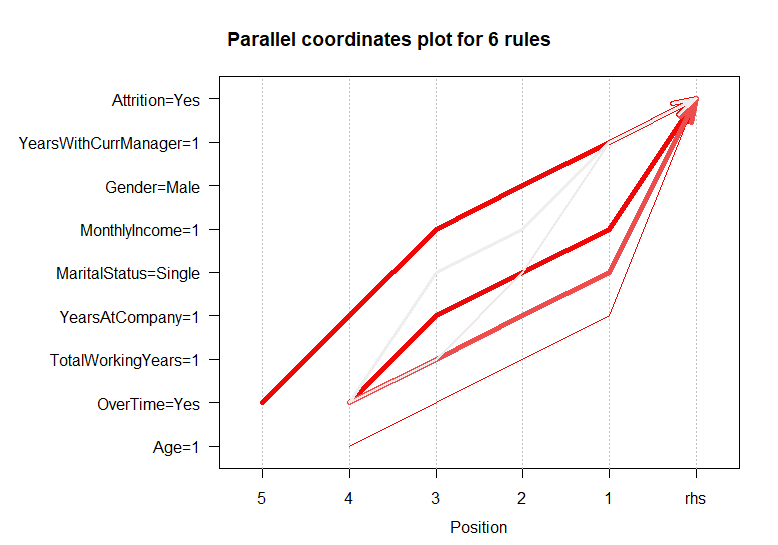
**Association**

Apriori Algorithm

We binned the continuous numeric variables and then converted them into factors and used it in the apriori algorithm along with the remaining categorical variables to find the association rules, specifically those having Attrition=Yes in the rhs.







Ranges for binned variables having binned value 1:

YearsWithCurrManager: 0-1 years

MonthlyIncome: 1009-4051

TotalWorkingYears: 0-6 years

YearsAtCompany: 0-2 years

Age: 18-25

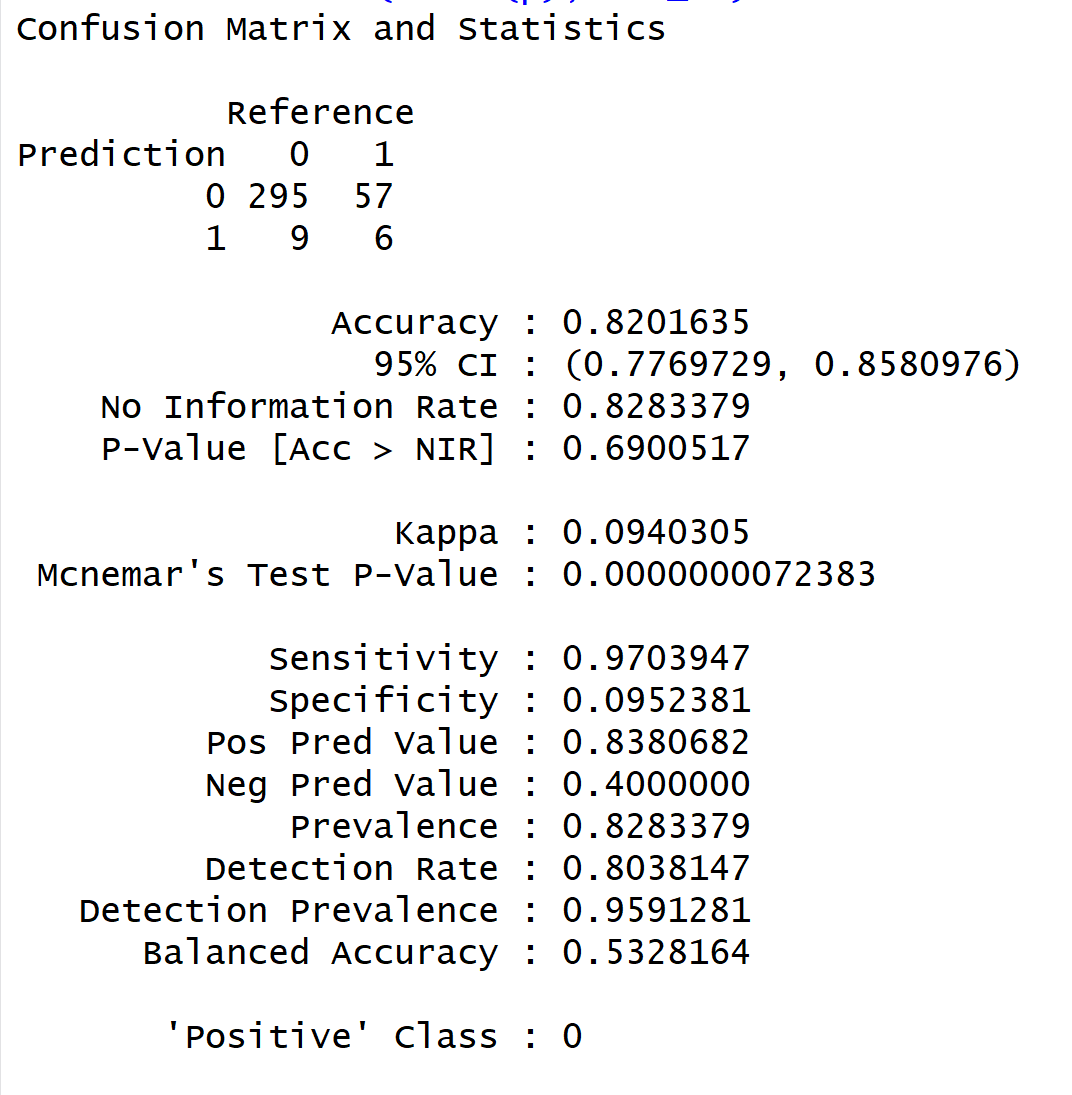
The association rules show that male employees who are single, work overtime, age in the range of 18-25 years & working for the company for around 0-2 years with monthly income in the range of $ 1009 - $ 4051 and working with the current manager for around 1 year show the tendency to leave the company.

**Evaluation**

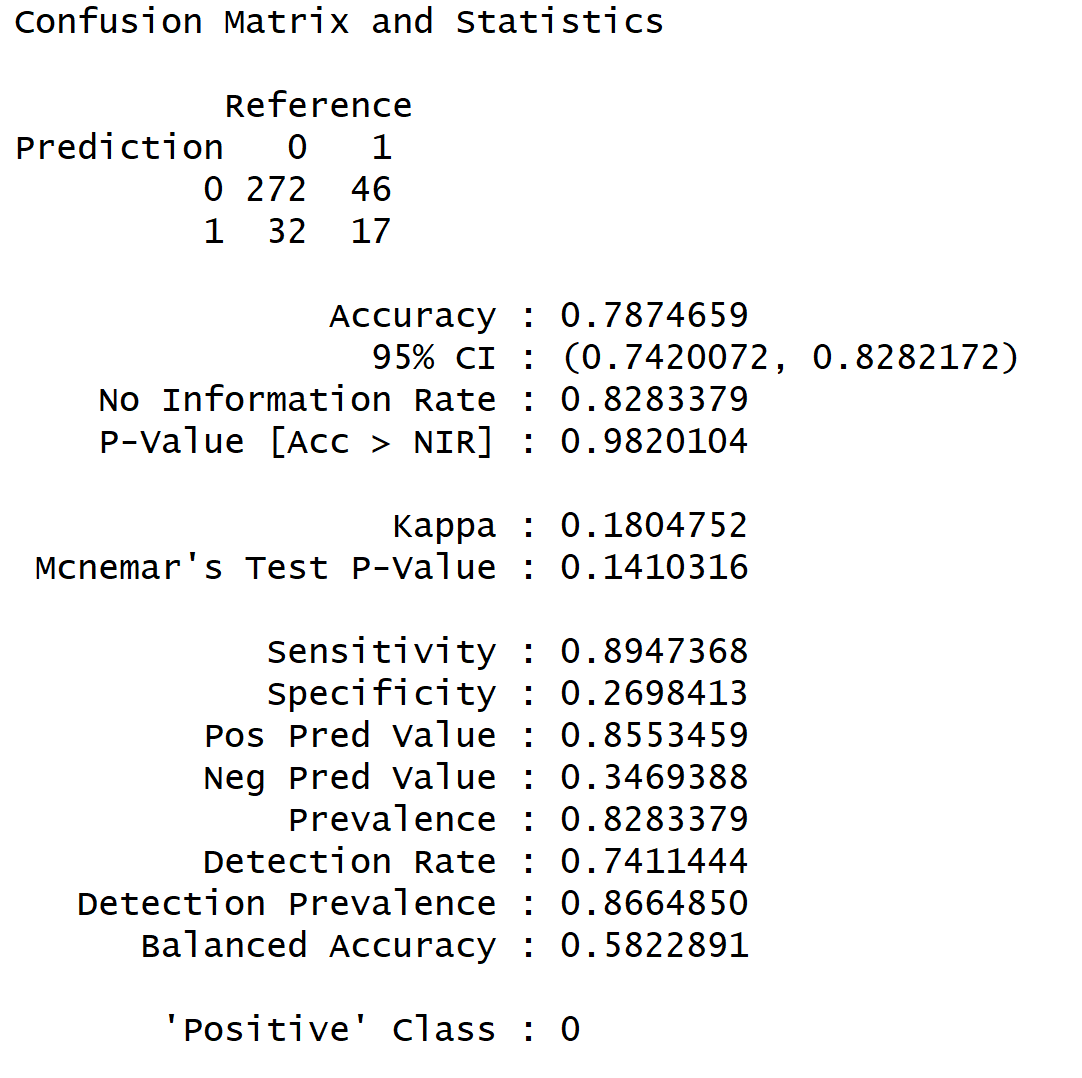
**Results**

The confusion matrix for the prediction algorithms used are as follows:

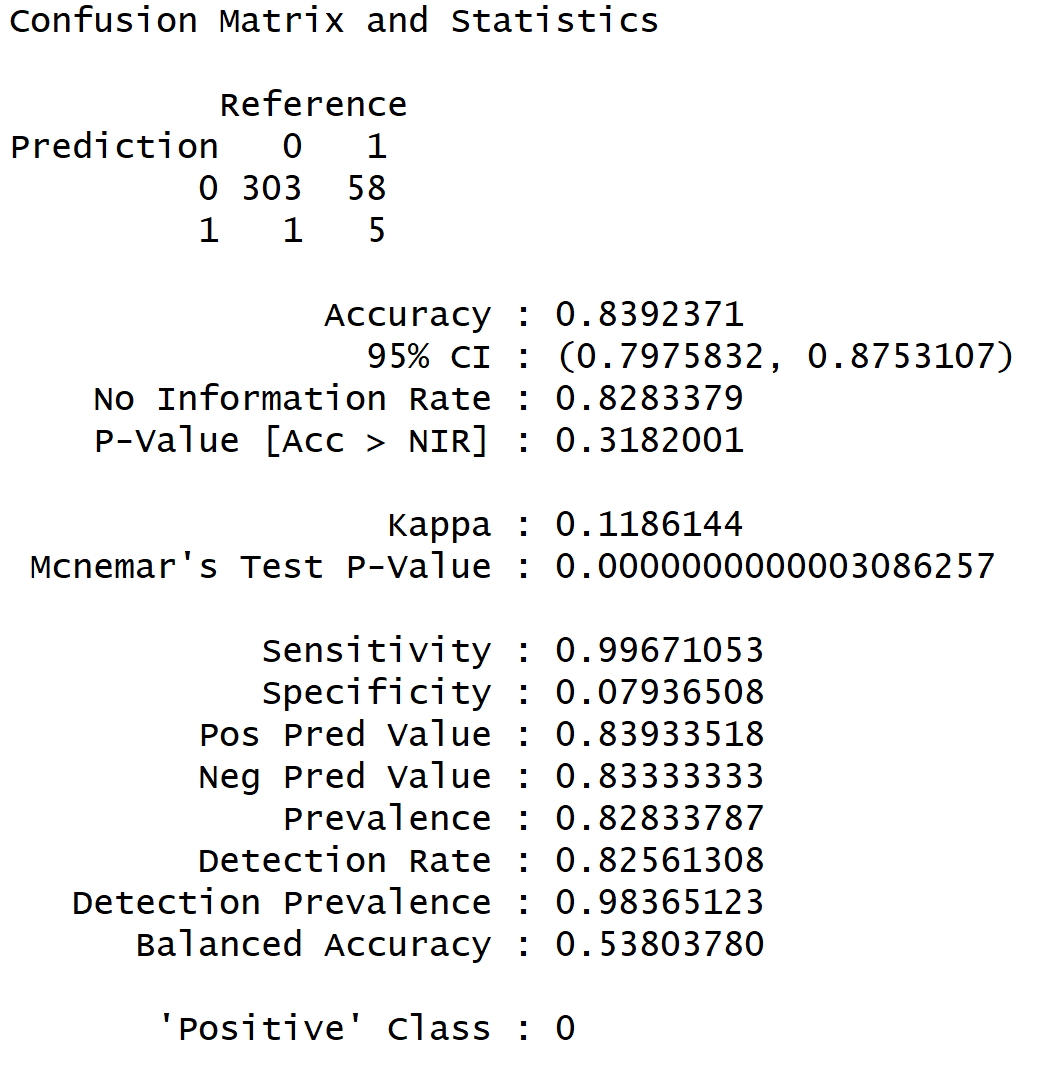
Classification and Regression Trees (CART)



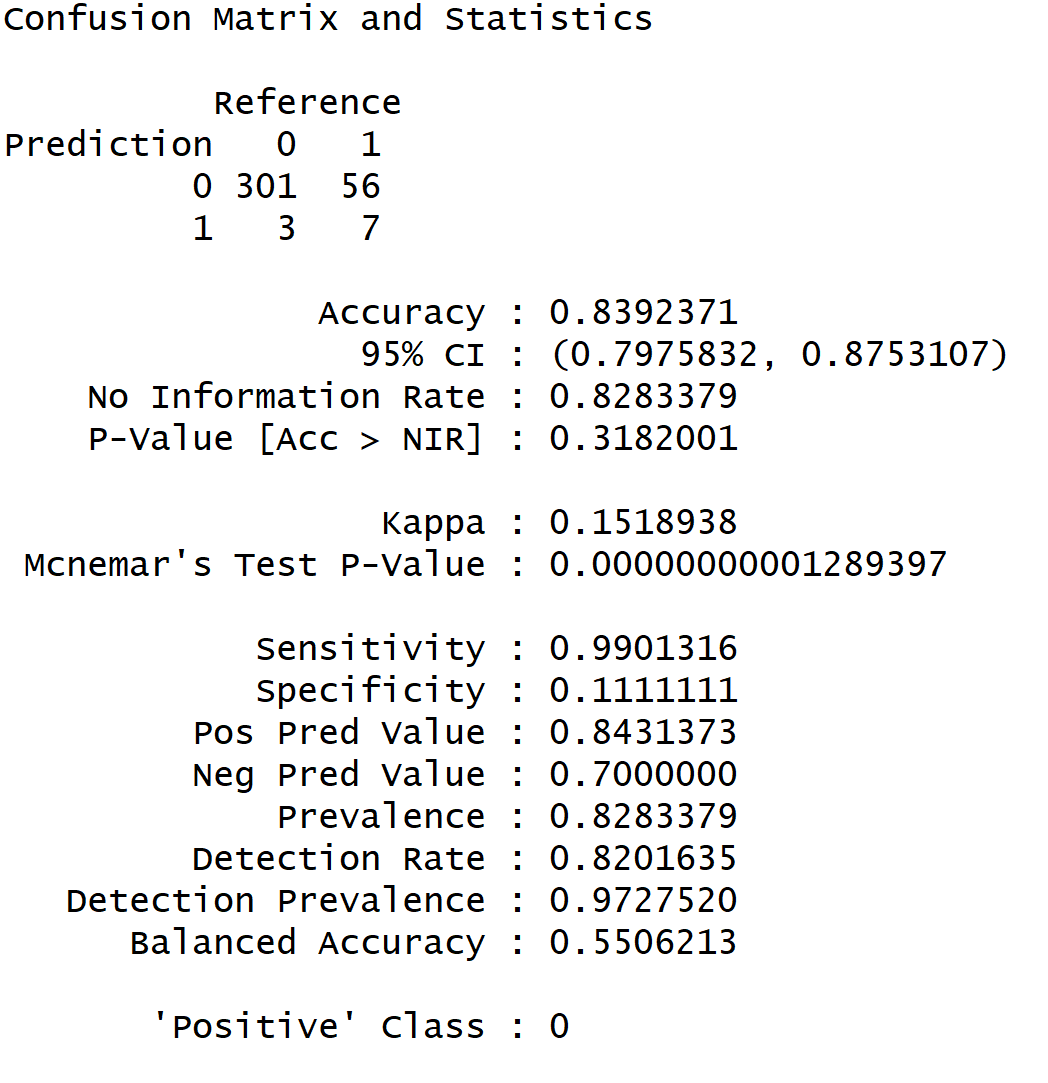
C5.0



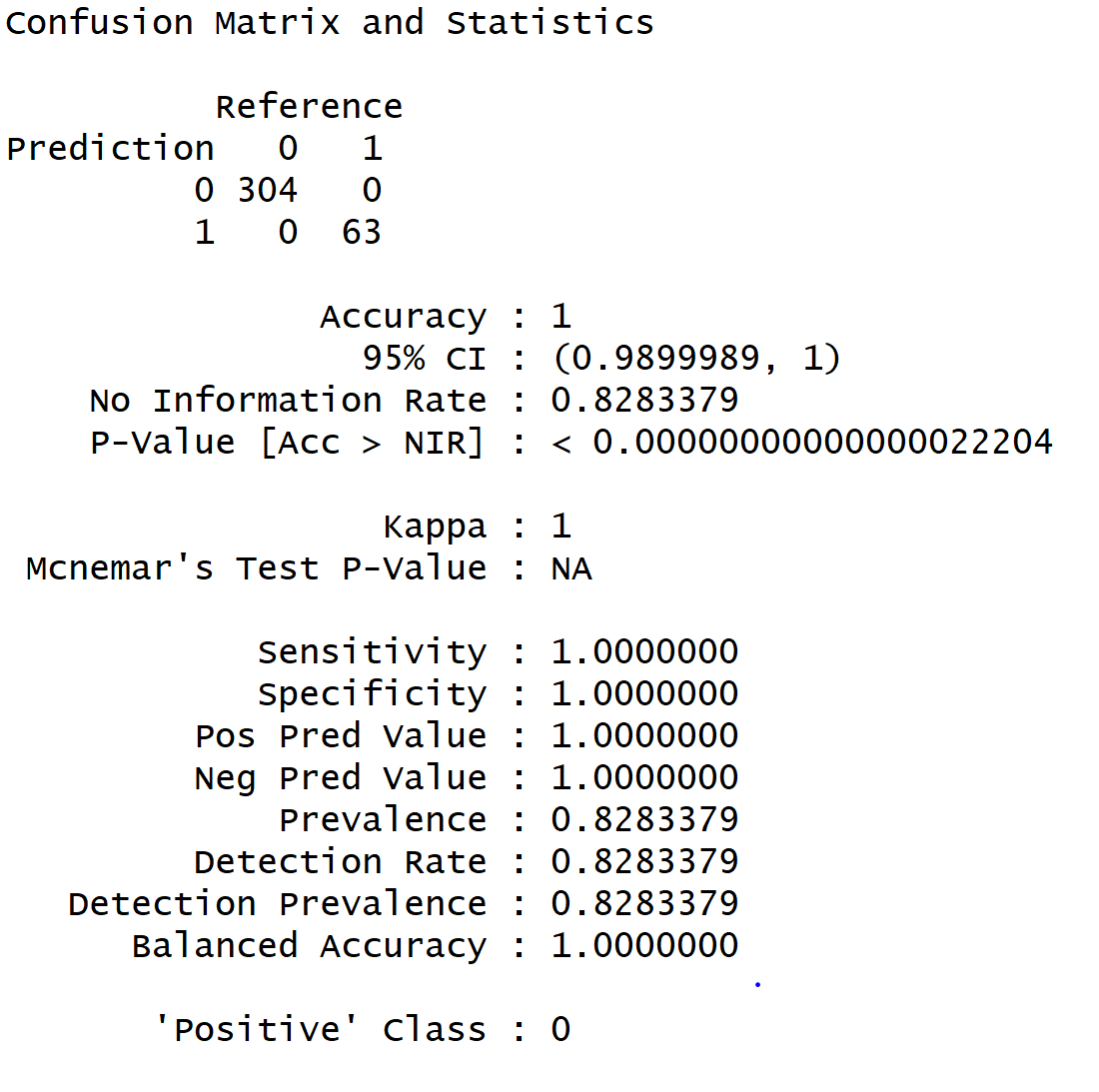
k-Nearest Neighbor Algorithm (KNN)



Random Forest



Neural Network



**Quality**

The accuracy rate for the prediction algorithms used are as follows:

|  |  |  |
| --- | --- | --- |
| **No.** | **Prediction Algorithm** | **Accuracy (%)** |
| 1. | Classification and Regression Trees (CART) | 82.01 |
| 2. | C5.0 | 78.74 |
| 3. | k-Nearest Neighbor Algorithm | 83.92 |
| 4. | Random Forest | 83.92 |
| 5. | Neural Network | 100 |

The output in the results section gives the details about sensitivity, specificity, prevalence, positive predictive value, negative predictive value, etc.

# **Report of Results**

**Knowledge discovered**

The data analysis using various techniques shows that the most important factors for the attrition are overtime, monthly salary, age, marital status, total working hours, number of years with current manager.

The association rule mining shows that male employees who are single, work overtime, age in the range of 18-25 years, working for the company for around 0-2 years with monthly income in the range of $ 1009 - $ 4051 and working with the current manager for around 1 year show the tendency to leave the company.

Cluster analysis shows that distance from home, Over time, training times last year, Work life balance plays an important role in Employee Attrition.

**Predictive Capabilities**

Neural Network has 100% prediction accuracy rate followed by Random Forest and KNN algorithms both having 83.92% accuracy rate.

**Limitations**

1. The dataset did not have any anomalous data or any missing data thus the exploratory data analysis was not much effective for this dataset.
2. Due to time constraint, we were not able to do correlation analysis.

**Future Work**

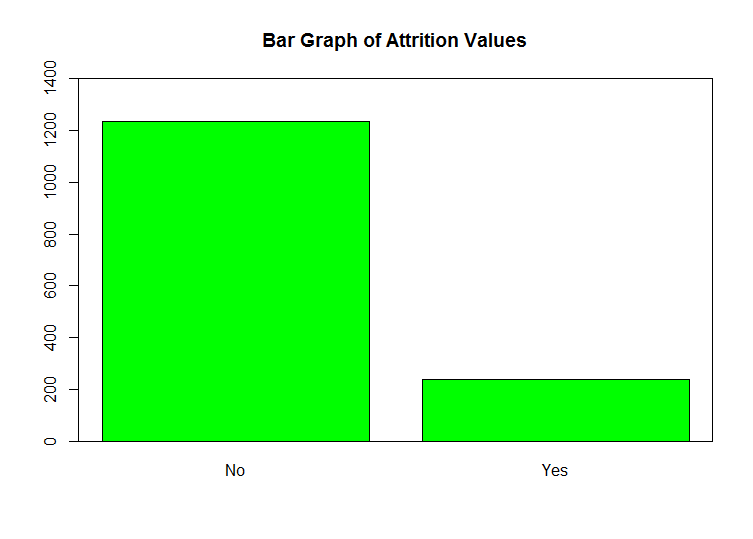
1. The future scope includes finding the correlation between various variables and performing correlation analysis to understand the 100% accuracy rate in case of neural networks modeling.
2. Finding the lift values for various models and plotting lift charts associated with them.
3. Use more algorithms like Support Vector Machines (SVM), Naive Bayes, Logistic Regression, Boosted Trees for modeling and evaluate the performance of the models.
4. Performing cluster analysis using k-means.

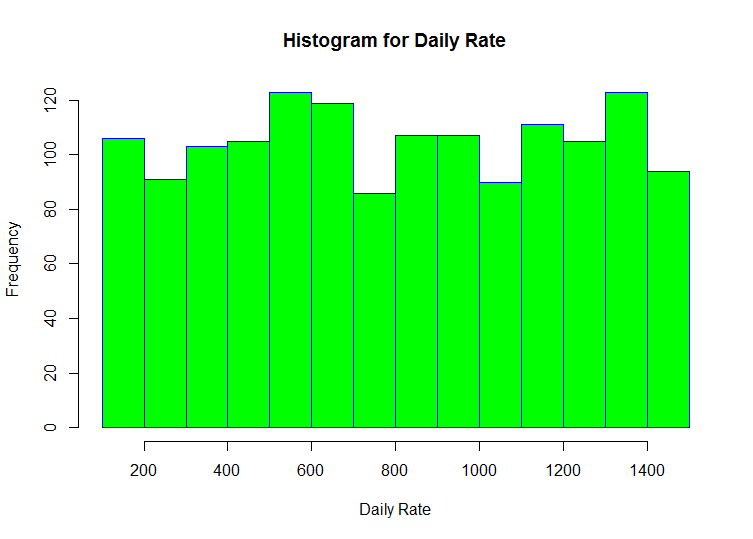
References

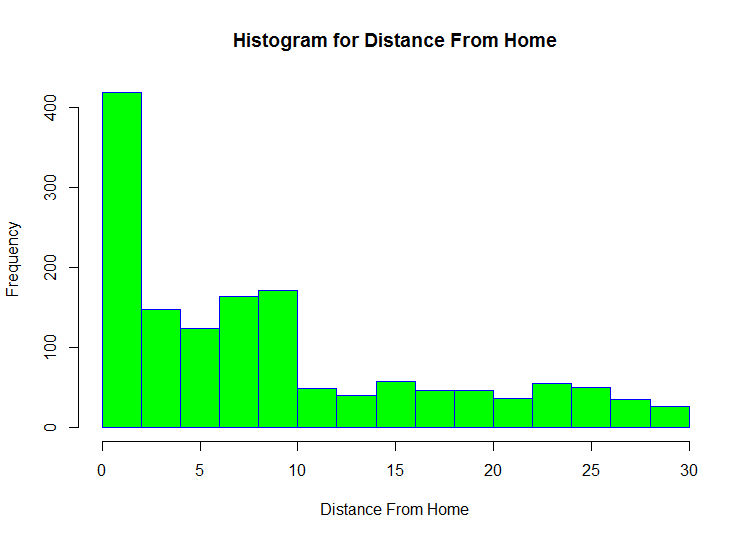
#### *Bhanuprakash*, (2017, April 21). IBM HR Data Analysis [Blog Post]. Retrieved from <https://rpubs.com/bpr1989/HRAnalysis>

*Mesfin Gebeyaw*, (2017, April 21). Using MCA and variable clustering in R for insights in customer attrition [Blog Post]. Retrieved from <https://datascienceplus.com/using-mca-and-variable-clustering-in-r-for-insights-in-customer-attrition/>

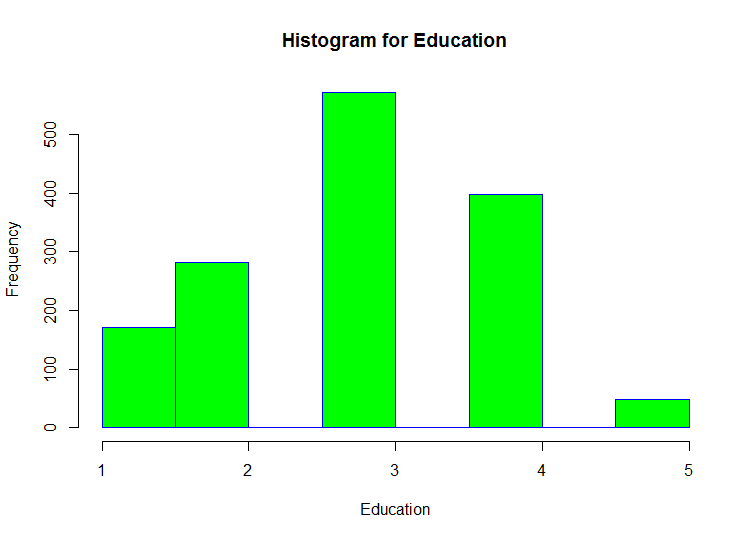
Appendix A



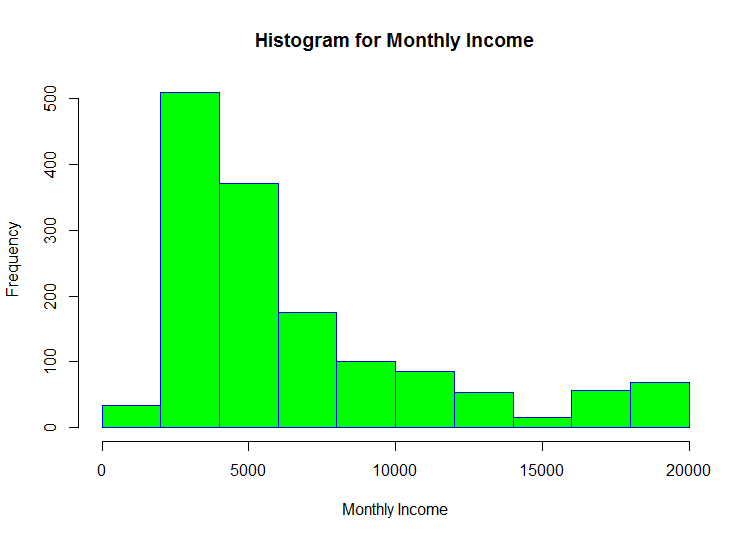


The daily rate seems to be a evenly distributed data

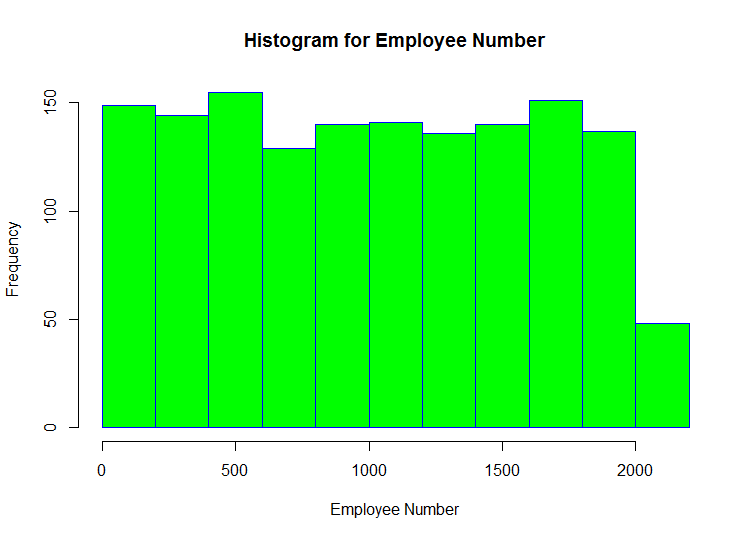
Majority employees stay in the vicinity of the company

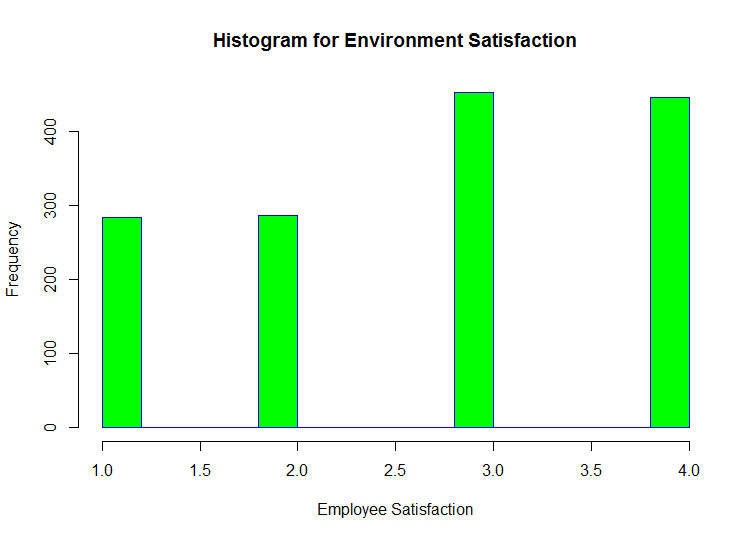


The histogram shows majority employees falling in the mid-educational bracket

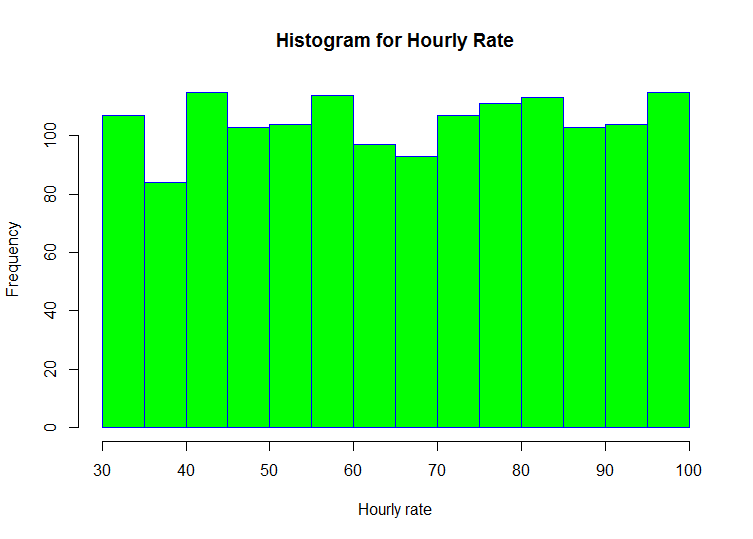


The histogram shows monthly income of the IBM employees



The histogram shows the employee number distribution Doesn’t look useful as of now but maybe important for other estimations.

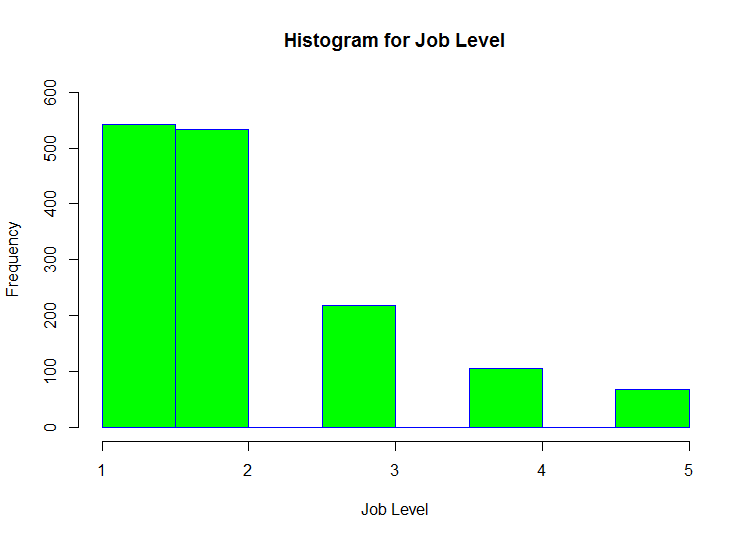
Observations tend to show that majority of employees are satisfied with the work



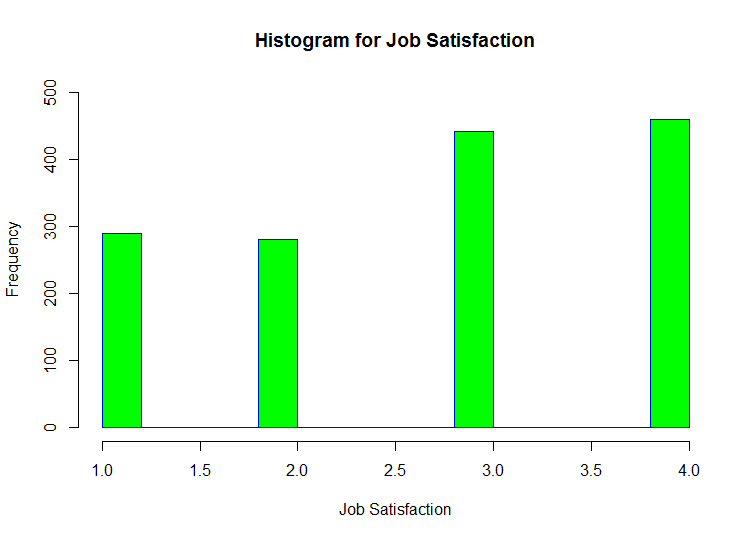
Observations shows evenly distributed/spread data (surprising!!!)



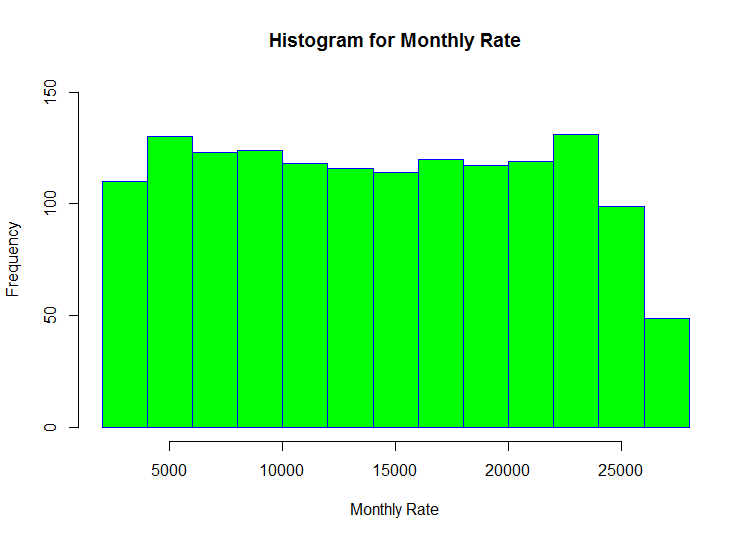
Histogram shows a decent level of work involvement

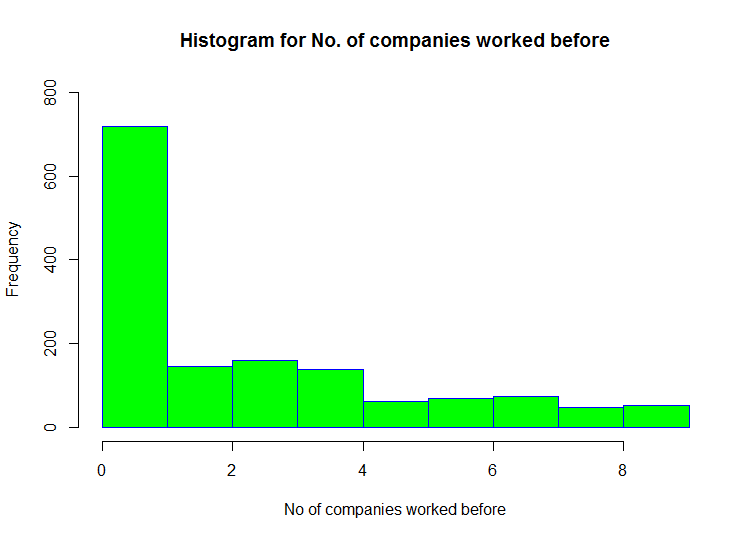


Histogram for job level vs number of employee at each job level

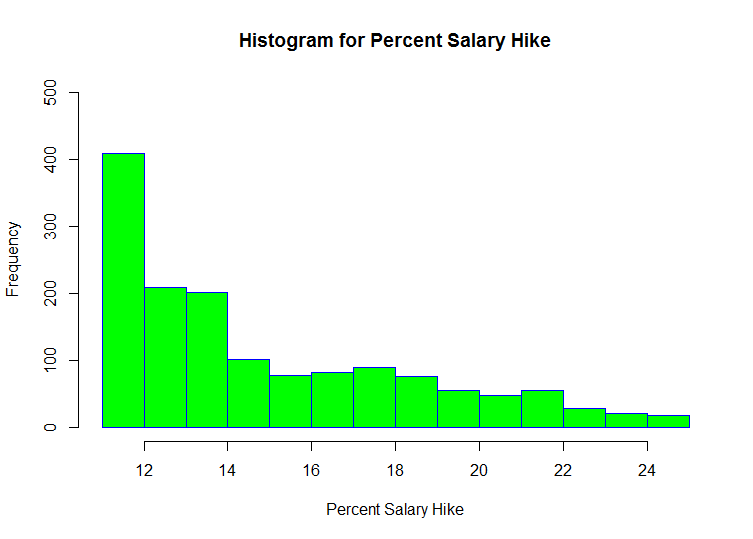


The histogram fairly gives us an idea that the employees are happy at IBM

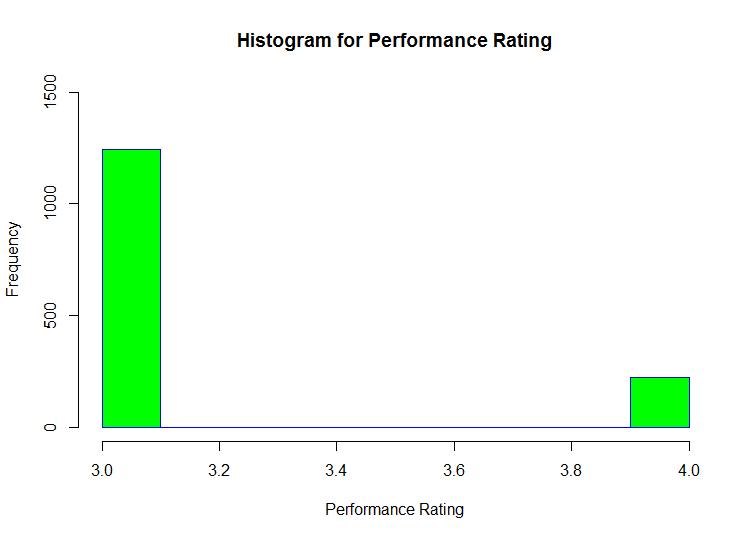
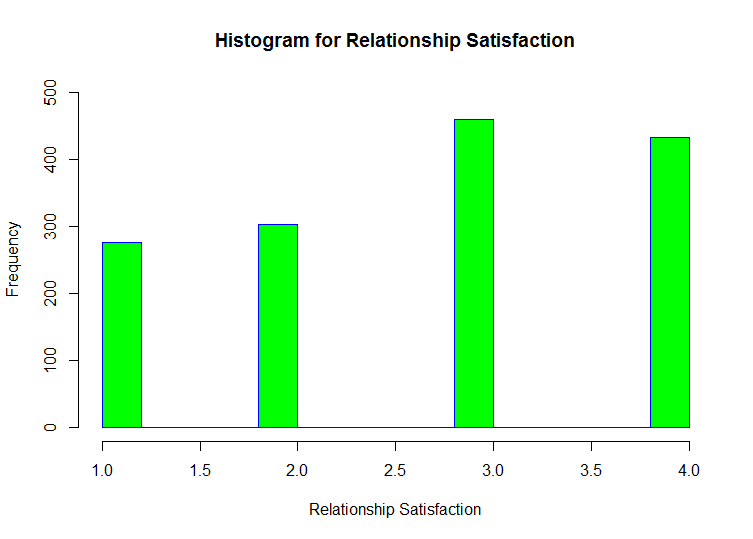




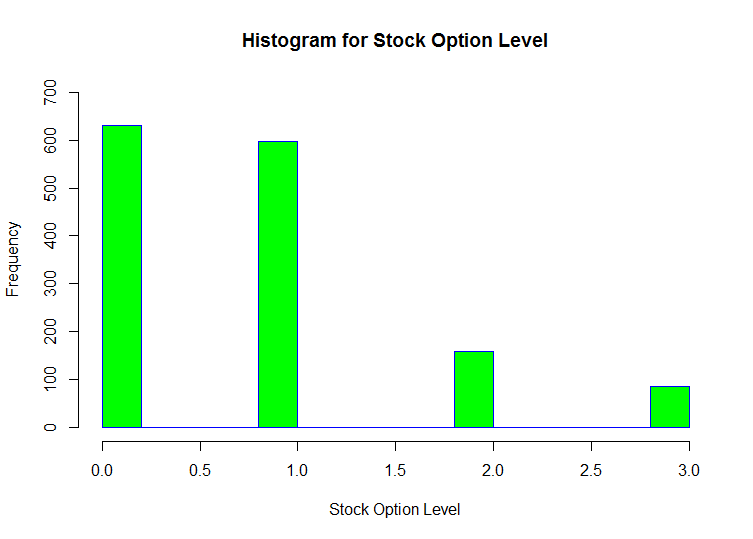
Histogram shows the number of companies IBM employees worked before.



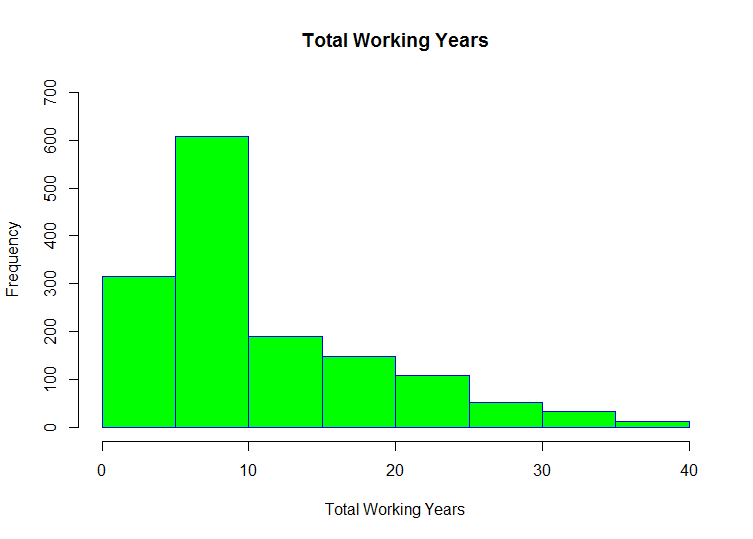
Histogram shows Percent salary hike for IBM employees



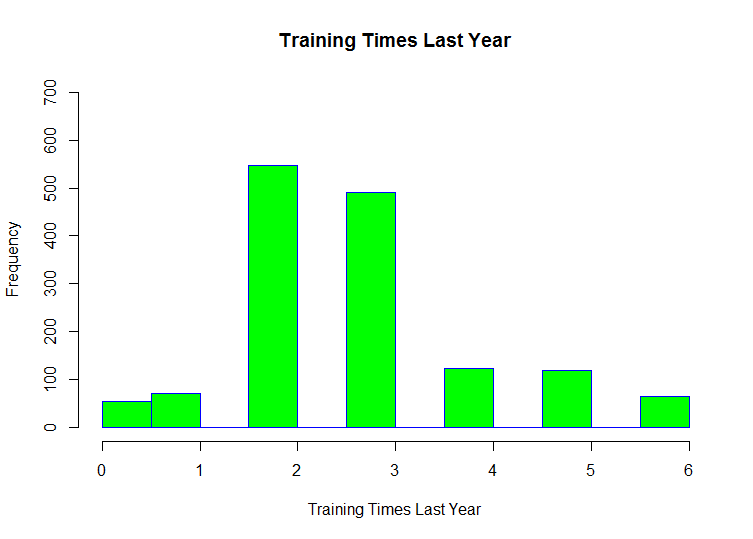
Histogram shows the Performance Rating of IBM Employees.



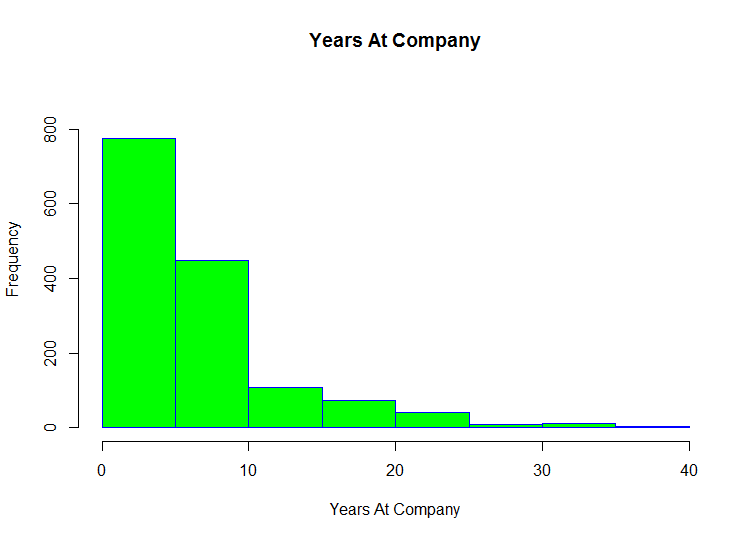
Histogram shows the Stock option level for IBM Employees.

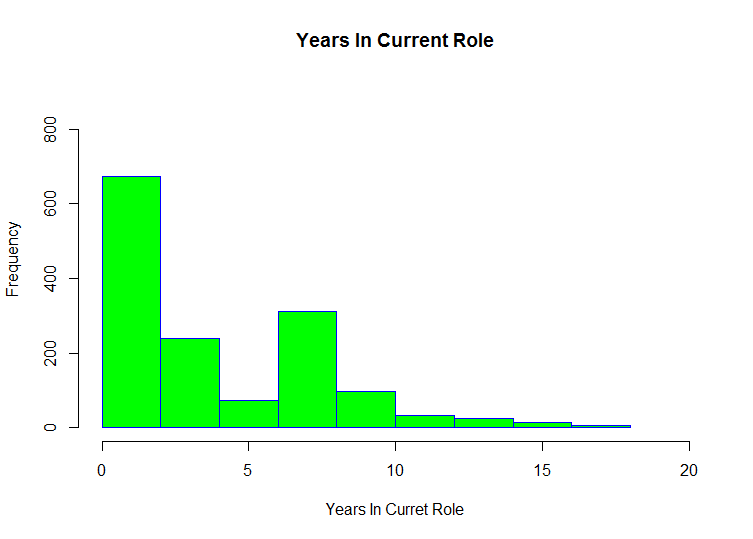


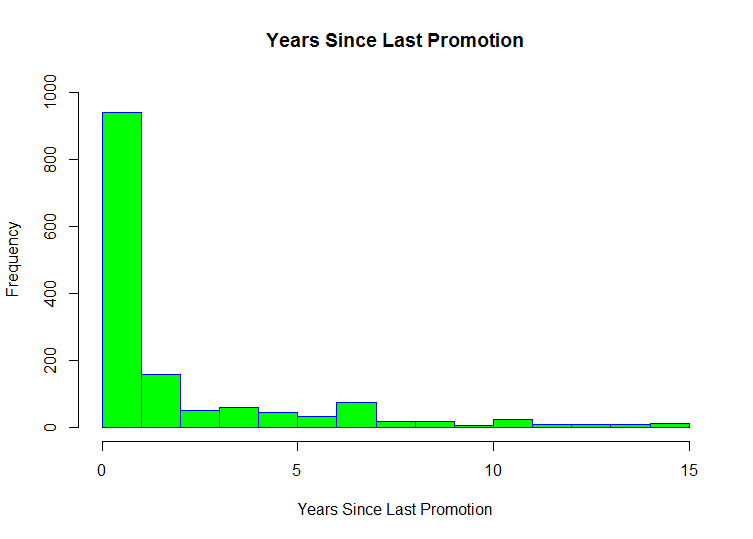
Histogram shows the Total Working Years for IBM Employees.

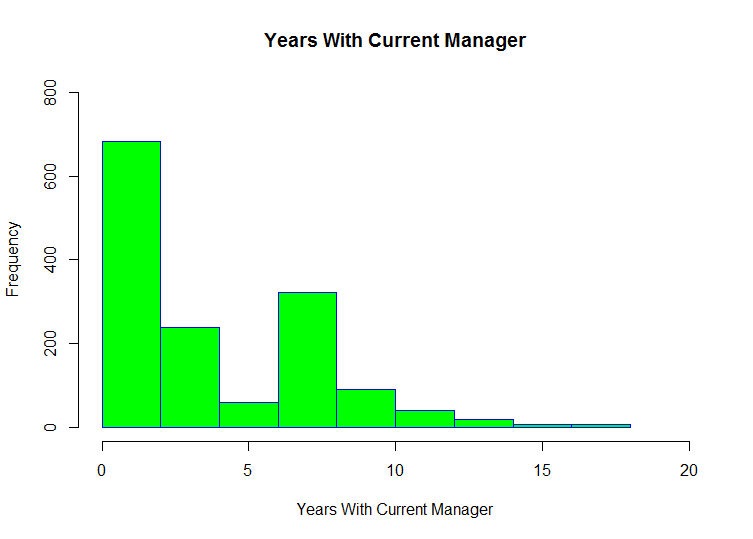












Appendix B

Let us calculate the summaries for each of the variables of the dataset

**Age**:

Min. 1st Qu. Median Mean 3rd Qu. Max.

18.00 30.00 36.00 36.92 43.00 60.00

**Attrition**:

Length Class Mode

1470 character character

**BusinessTravel**

Length Class Mode

1470 character character

**DailyRate**

Min. 1st Qu. Median Mean 3rd Qu. Max.

102.0 465.0 802.0 802.5 1157.0 1499.0

**Department**

Length Class Mode

1470 character character

**DistanceFromHome**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.000 7.000 9.193 14.000 29.000

**Education**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.000 3.000 2.913 4.000 5.000

**EducationField**

Length Class Mode

1470 character character

**EnvironmentSatisfaction**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.000 3.000 2.722 4.000 4.000

**Gender**

Length Class Mode

1470 character character

**HourlyRate**

Min. 1st Qu. Median Mean 3rd Qu. Max.

30.00 48.00 66.00 65.89 83.75 100.00

**JobInvolvement**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.00 2.00 3.00 2.73 3.00 4.00

**JobLevel**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 1.000 2.000 2.064 3.000 5.000

**JobRole**

Length Class Mode

1470 character character

**JobSatisfaction**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.000 3.000 2.729 4.000 4.000

**MaritalStatus**

Length Class Mode

1470 character character

**MonthlyIncome**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1009 2911 4919 6503 8379 19999

**MonthlyRate**

Min. 1st Qu. Median Mean 3rd Qu. Max.

2094 8047 14236 14313 20462 26999

**NumCompaniesWorked**

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.000 1.000 2.000 2.693 4.000 9.000

**Over18**

Length Class Mode

1470 character character

**OverTime**

Length Class Mode

1470 character character

**PercentSalaryHike**

Min. 1st Qu. Median Mean 3rd Qu. Max.

11.00 12.00 14.00 15.21 18.00 25.00

**PerformanceRating**

Min. 1st Qu. Median Mean 3rd Qu. Max.

3.000 3.000 3.000 3.154 3.000 4.000

**RelationshipSatisfaction**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.000 3.000 2.712 4.000 4.000

**StandardHours**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.000 3.000 2.712 4.000 4.000

**StockOptionLevel**

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.0000 0.0000 1.0000 0.7939 1.0000 3.0000

**TotalWorkingYears**

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.00 6.00 10.00 11.28 15.00 40.00

**TrainingTimesLastYear**

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.000 2.000 3.000 2.799 3.000 6.000

**WorkLifeBalance**

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.000 3.000 2.761 3.000 4.000

**YearsAtCompany**

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.000 3.000 5.000 7.008 9.000 40.000

**YearsInCurrentRole**

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.000 2.000 3.000 4.229 7.000 18.000

**YearsSinceLastPromotion**

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.000 0.000 1.000 2.188 3.000 15.000

**YearsWithCurrManager**

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.000 2.000 3.000 4.123 7.000 17.000