Statistical Analysis and Prediction of IBM Employee Attrition

Meenakshi Rajgopal

Contents

Introduction

Data set is taken from Kaggle.com -https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset. It has several variables which can help us to understand which of these are causing the increase of attrition at IBM. There is a total of 1470 observations and 13 variables. R studio will be used for the analysis and MicroStrategy desktop for the visual charts. This analysis of IBM dataset is to understand what are the factors that are contributing to the attrition of employees in the Organization.

Literature Review

Background of attrition in the organizations.

Attrition has always been around affecting any organizations in the industry. However, over the past few years' companies have taken more interest in looking at their HR practices on how it aligns with their business. Any attrition due to resignation, retirement or death, etc. organizations are still trying to understand the best practices and ways to reduce it. Dobhal and Nigam (2018) have attempted to understand this through their research paper on Employee Attrition and Employee satisfaction. The authors describe that the reason behind these attritions can be looked at statistically by the organizational variables like occupational category, location, benefits, promotion, payment and many more. There can also be some demographic variables which influence attrition like working condition, job satisfaction, etc. After performing correlation, Hypothesis tests it can be given a much clear picture for HR allowing them to work with the management on the execution plans.

The author also emphasizes that every organization will have to work on their motivational system based on the findings from the analysis. Plonski (2017) has explained how machine learning can assist in predicting employee attrition. He describes how it can be financially beneficial for the organization if the management knew when the employee might quit. There are costs involved in training these employees and if the HR and management can make strategies on how a small amount of retention bonus can help to retain a good hire. Through his studies, the author could predict business travel was one of the variables during the decision-making process in the attrition. It can be different for different organizations.

Financial Impact with Employee Turnover

In the research paper titled “Retaining talent”, Allen (n.d.) has spoken about employee turnover. He has focused more on why retaining the talent and turnover important for the organization and how it can be achieved through statistical analysis. He asserts that every employee departure costs the organization time, money and other resources. Employee turnover can be voluntary as well as involuntary. As per the US Bureau of Labor Statistics, 23.7% of American workers voluntarily quit their jobs in 2006. That is quite an impactful percentage which could be addressed. The author has listed different categories of turnover predictors. They range from stronger to weaker relationship to turnover. These variables can vary depending on job types and companies. He has listed the ways to improve this turnover rate example making more out of exit interviews to understand the true reason, post-exit anonymous surveys, etc. which will help after we know the reasons behind employees leaving the organization. Frye (2018) dig deeper into the employee attrition and tried to understand more about the financial losses by correlation and Principal component analysis. In this analysis, attrition is highly correlated with length of service. One other thing that correlates the most is salary and promotion. The principal component analysis allowed him to reduce from the initial 99 components to 46 most important components which should be used in the model. This is done to ensure they have the correct data to do analysis and not include unnecessary data.

Studies made in Center's for American Progress analysis (2012) indicate the separation of employees can cost somewhere between 16% to 213% of the annual salary depending on the job type. The author further described that after their initial technique to simplify the data, they only focused on the variables that were within employers' control and others would be considered to narrow down the possibilities. Sen Gupta, (n.d.) examines the attrition at the BPO industry and the variables affecting this industry.

The author goes on to talk more about her research to produce a model for employee retention adjoining it with other aspects of apparent attitudes like employee motivation, employee satisfaction, employee involvement, and life interest and work compatibility, etc. Many researchers have extensively worked on this topic to find out the combination of factors to improve the attrition rate. The article "Employee Attrition and Retention" has conducted multiple statistical analysis like regression, t-tests, and correlation. The principal component analysis was used for extraction. Correlation explains job nature, personal factors, hostile organizational support are the main determinants of attrition in the BPO industry. These dimensions were further compared with personal characteristics like marital status, education, and age to get a broader picture. For education, the career path has significant differences in mean and standard deviation. For marital status and age monotonous job, unfavorable working conditions, hostile environment were found to have significant differences. The final regression analysis was done with the dependent variable as attrition or sustenance and independent variables are age, gender, education, marital status, etc. Based on this analysis author concluded that motivational factors had the most significance when compared by actual and predicted values. The correlation showed 0.659 at a significant level of 0.01.

Conerly (2018) focused more on employee turnover. He explains that key to improving the employee attrition is understanding the turnover rate. Before engaging the employees on small or big programs it is important to know its worth and how much will it benefit. It is not only an issue with HR but even operations feel the heat in terms of loss of talent, reduced customer service and reduced productivity. Conerly suggests that a team comprising of HR, operations and finance would be the best possible team to contribute towards reducing the turnover rate. The costs incurring right from HR exit interview to employing a new person, orientation, training, etc. which are the main HR costs. There are other costs like existing employees covering for the person who leaves, their overtime, loss of productivity is outside HR but still an incurring cost. Anderson (2018) states that it is a universally known fact that employees who are happy at their work stay for a longer duration. However, which are these factors that are making employees happy or have a positive correlation with retention. According to the author after looking at several pieces of research the factors that look more influenced by age are technology, communication, and adaptability. A very important suggestion given by the author is that, since baby boomers are approaching retirement suggestions from them will be valuable for the HR and management to work together. The collection of reviews data available at job portals like Glassdoor, ZipRecruiter should be included. Also, crowdsourcing opinions from people will help understand their perception from outside. Anderson also emphasized that 235,000 exit interviews were analyzed by researchers to understand that the main reason for employees leaving was lack of opportunity, work-life balance and manager behavior.

Business Problem

Find the factors which are increasing the employee attrition rate and provide a recommendation to alleviate these factors. How can these factors be addressed to retain the talent and reduce financial impact after investing in these employees?

Dataset Description

The section below will show the variables included in the data set and their basic statistical summary by reading the CSV file. This dataset will be used for our further analysis.

Some of the libraries will be required for this analysis.

```
library("RColorBrewer")
library(corrplot)
library(rpart)
library(rpart.plot)
```

Figure 1 Reading the dataset and look at the variables.

R's default behavior is to convert all the characters in factors because of which character columns does not give the result we are looking for. To avoid this problem, the data is read as stringsAsfactors = False.

```
> ibmattrition <- read.csv("Attrition.csv", header = T, stringsAsFactors = FALSE)
> names(ibmattrition)
 [1] "ï..Age"                  "Attrition"                "BusinessTravel"
 [4] "DailyRate"               "Department"               "DistanceFromHome"
 [7] "Education"               "EducationField"           "EmployeeCount"
[10] "EmployeeNumber"          "EnvironmentSatisfaction"  "Gender"
[13] "HourlyRate"              "JobInvolvement"           "JobLevel"
[16] "JobRole"                 "JobSatisfaction"          "MaritalStatus"
[19] "MonthlyIncome"           "MonthlyRate"              "NumCompaniesWorked"
[22] "Over18"                  "OverTime"                 "PercentSalaryHike"
[25] "PerformanceRating"       "RelationshipSatisfaction" "StandardHours"
[28] "StockOptionLevel"        "TotalWorkingYears"        "TrainingTimesLastYear"
[31] "WorkLifeBalance"         "YearsAtCompany"           "YearsInCurrentRole"
[34] "YearsSinceLastPromotion" "YearsWithCurrManager"
```

Figure 2 Statistical summary.

```
summary(ibmattrition)
     ï..Age      Attrition          BusinessTravel       DailyRate       Department
 Min.   :18.00   Length:1470        Length:1470        Min.   : 102.0   Length:1470
 1st Qu.:30.00   Class :character   Class :character   1st Qu.: 465.0   Class :character
 Median :36.00   Mode  :character   Mode  :character   Median : 802.0   Mode  :character
 Mean   :36.92                                         Mean   : 802.5
 3rd Qu.:43.00                                         3rd Qu.:1157.0
 Max.   :60.00                                         Max.   :1499.0
 DistanceFromHome    Education     EducationField     EmployeeCount EmployeeNumber
 Min.   : 1.000   Min.   :1.000   Length:1470        Min.   :1     Min.   :   1.0
 1st Qu.: 2.000   1st Qu.:2.000   Class :character   1st Qu.:1     1st Qu.: 491.2
 Median : 7.000   Median :3.000   Mode  :character   Median :1     Median :1020.5
 Mean   : 9.193   Mean   :2.913                      Mean   :1     Mean   :1024.9
 3rd Qu.:14.000   3rd Qu.:4.000                      3rd Qu.:1     3rd Qu.:1555.8
 Max.   :29.000   Max.   :5.000                      Max.   :1     Max.   :2068.0
 EnvironmentSatisfaction    Gender            HourlyRate     JobInvolvement    JobLevel
 Min.   :1.000           Length:1470        Min.   : 30.00   Min.   :1.00   Min.   :1.000
 1st Qu.:2.000           Class :character   1st Qu.: 48.00   1st Qu.:2.00   1st Qu.:1.000
 Median :3.000           Mode  :character   Median : 66.00   Median :3.00   Median :2.000
 Mean   :2.722                              Mean   : 65.89   Mean   :2.73   Mean   :2.064
 3rd Qu.:4.000                              3rd Qu.: 83.75   3rd Qu.:3.00   3rd Qu.:3.000
 Max.   :4.000                              Max.   :100.00   Max.   :4.00   Max.   :5.000
   JobRole          JobSatisfaction MaritalStatus      MonthlyIncome    MonthlyRate
 Length:1470        Min.   :1.000   Length:1470        Min.   : 1009   Min.   : 2094
 Class :character   1st Qu.:2.000   Class :character   1st Qu.: 2911   1st Qu.: 8047
 Mode  :character   Median :3.000   Mode  :character   Median : 4919   Median :14236
                    Mean   :2.729                      Mean   : 6503   Mean   :14313
                    3rd Qu.:4.000                      3rd Qu.: 8379   3rd Qu.:20462
                    Max.   :4.000                      Max.   :19999   Max.   :26999
 NumCompaniesWorked    Over18             OverTime         PercentSalaryHike PerformanceRating
 Min.   :0.000      Length:1470        Length:1470        Min.   :11.00     Min.   :3.000
 1st Qu.:1.000      Class :character   Class :character   1st Qu.:12.00     1st Qu.:3.000
 Median :2.000      Mode  :character   Mode  :character   Median :14.00     Median :3.000
 Mean   :2.693                                            Mean   :15.21     Mean   :3.154
 3rd Qu.:4.000                                            3rd Qu.:18.00     3rd Qu.:3.000
 Max.   :9.000                                            Max.   :25.00     Max.   :4.000
 RelationshipSatisfaction StandardHours StockOptionLevel TotalWorkingYears TrainingTimesLastYear
 Min.   :1.000            Min.   :80    Min.   :0.0000   Min.   : 0.00     Min.   :0.000
 1st Qu.:2.000            1st Qu.:80    1st Qu.:0.0000   1st Qu.: 6.00     1st Qu.:2.000
 Median :3.000            Median :80    Median :1.0000   Median :10.00     Median :3.000
 Mean   :2.712            Mean   :80    Mean   :0.7939   Mean   :11.28     Mean   :2.799
 3rd Qu.:4.000            3rd Qu.:80    3rd Qu.:1.0000   3rd Qu.:15.00     3rd Qu.:3.000
 Max.   :4.000            Max.   :80    Max.   :3.0000   Max.   :40.00     Max.   :6.000
 WorkLifeBalance YearsAtCompany   YearsInCurrentRole YearsSinceLastPromotion YearsWithCurrManager
 Min.   :1.000   Min.   : 0.000   Min.   : 0.000     Min.   : 0.000          Min.   : 0.000
 1st Qu.:2.000   1st Qu.: 3.000   1st Qu.: 2.000     1st Qu.: 0.000          1st Qu.: 2.000
 Median :3.000   Median : 5.000   Median : 3.000     Median : 1.000          Median : 3.000
 Mean   :2.761   Mean   : 7.008   Mean   : 4.229     Mean   : 2.188          Mean   : 4.123
 3rd Qu.:3.000   3rd Qu.: 9.000   3rd Qu.: 7.000     3rd Qu.: 3.000          3rd Qu.: 7.000
 Max.   :4.000   Max.   :40.000   Max.   :18.000     Max.   :15.000          Max.   :17.000
```

End of docum

Data Exploration

This section will allow us to identify the variables involved in this dataset. In this dataset, we have some variables which are characters and we need numeric to run the correlation matrix. Converting these variables to numeric.

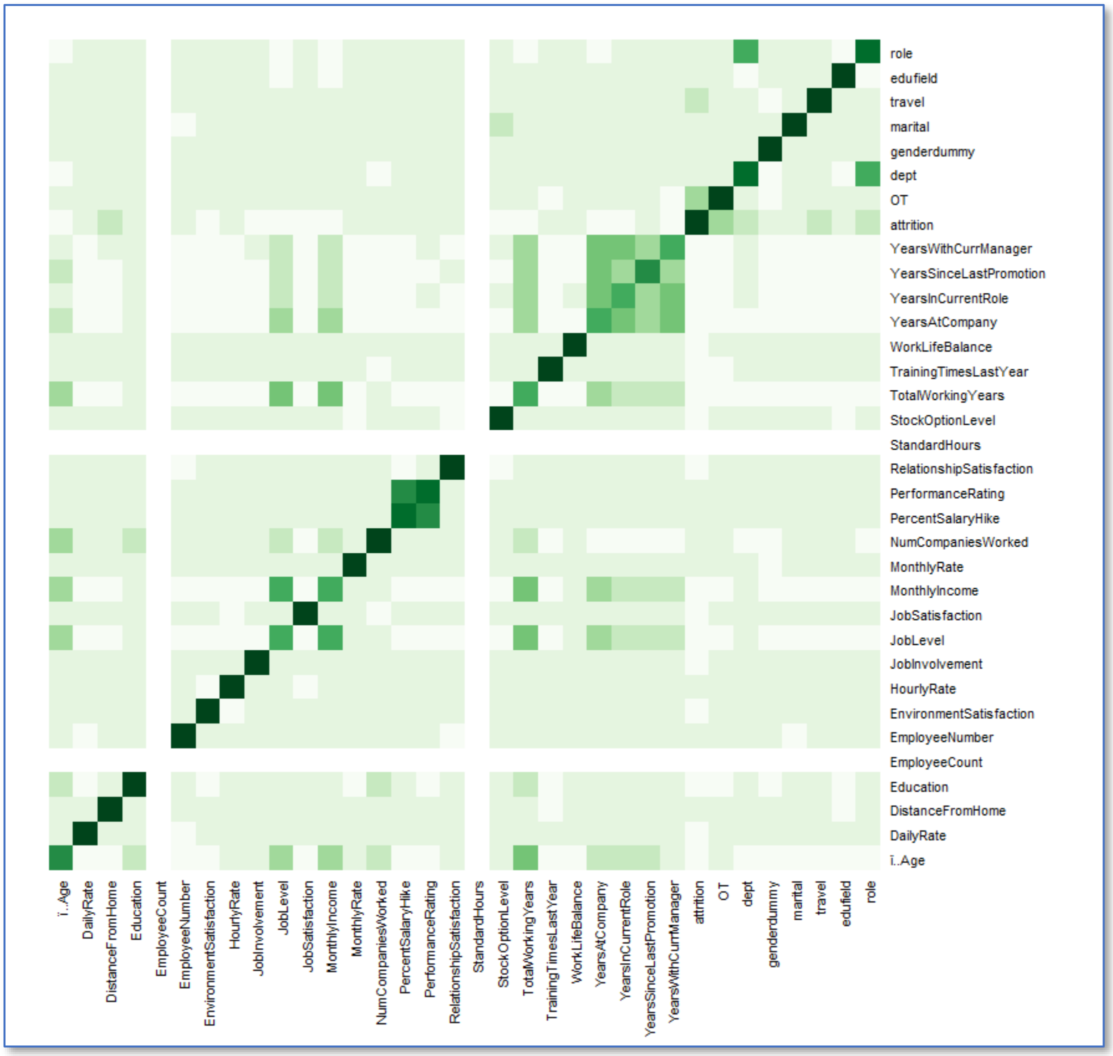
Figure 3 Converting to numeric.

```
> ##Converting to numeric
> ibmattrition$attrition <- ifelse(ibmattrition$Attrition == "Yes",1,0)
>
> ibmattrition$OT <- ifelse(ibmattrition$OverTime == "Yes",1,0)
>
> ibmattrition$dept <- ifelse(ibmattrition$Department == "Human Resources",0,
+                             ifelse(ibmattrition$Department == "Research & Development",1,2))
>
> ibmattrition$genderdummy <- ifelse(ibmattrition$Gender == "Female",0,1)
>
> ibmattrition$marital <- ifelse(ibmattrition$MaritalStatus == "Single",1,
+                                ifelse(ibmattrition$MaritalStatus == "Married",0,2))
>
> ibmattrition$travel <- ifelse(ibmattrition$BusinessTravel == "Non-Travel",0,
+                               ifelse(ibmattrition$BusinessTravel == "Travel_Rarely",1,2))
>
> ibmattrition$edufield <- ifelse(ibmattrition$EducationField == "Human Resources",0,
+                               ifelse(ibmattrition$BusinessTravel == "Life Sciences",1,
+                                      ifelse(ibmattrition$EducationField == "Marketing",2,
+                                             ifelse(ibmattrition$EducationField == "Medical",3,
+                                                    ifelse(ibmattrition$EducationField == "Other",4,5)))))
>
> ibmattrition$role <- ifelse(ibmattrition$JobRole == "Healthcare Representative",0,
+                             ifelse(ibmattrition$JobRole == "Human Resources",1,
+                                    ifelse(ibmattrition$JobRole == "Laboratory Technician",2,
+                                           ifelse(ibmattrition$JobRole == "Manager",3,
+                                                  ifelse(ibmattrition$JobRole == "Manufacturing Director",4,
+                                                         ifelse(ibmattrition$JobRole == "Research Director",5,
+                                                                ifelse(ibmattrition$JobRole == "Research Scientist",6,
+                                                                       ifelse(ibmattrition$JobRole == "Sales Executive",7,8))))))))
>
```

Now running the correlation matrix.

Figure 4 Heatmap

```
## HeatMap
heatmap(cor(ibmattrition[sapply(ibmattrition, is.numeric)]),
        Rowv = NA, Colv = NA,col=brewer.pal(9,"Greens"),margins=c(9,9))
```

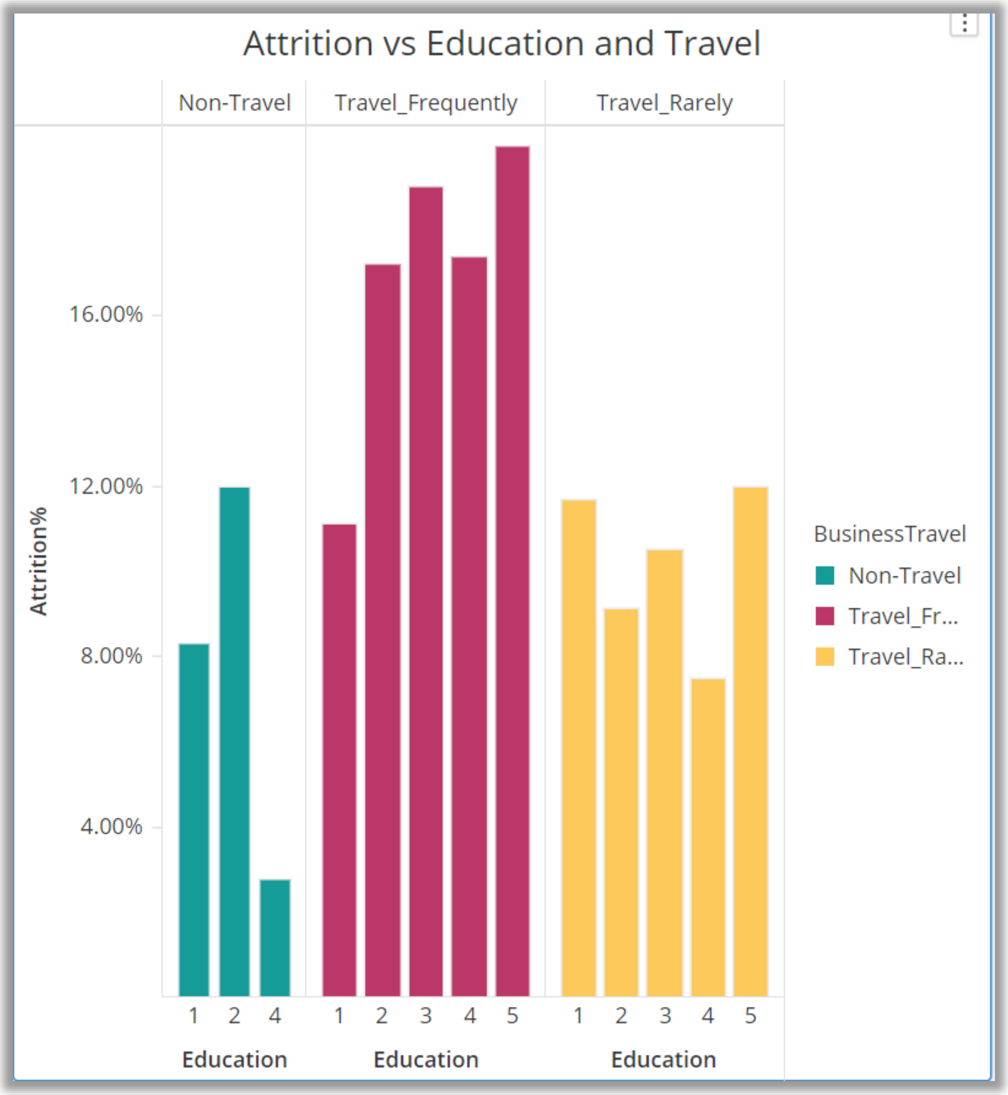


The correlation matrix indicates variable Attrition has a positive correlation with Overtime and Travel. It also has a negative correlation with Years with current manager, years at the company, years since promotion and years at the current role. We will use these variables in the decision tree to predict and decide on what is causing these attritions.

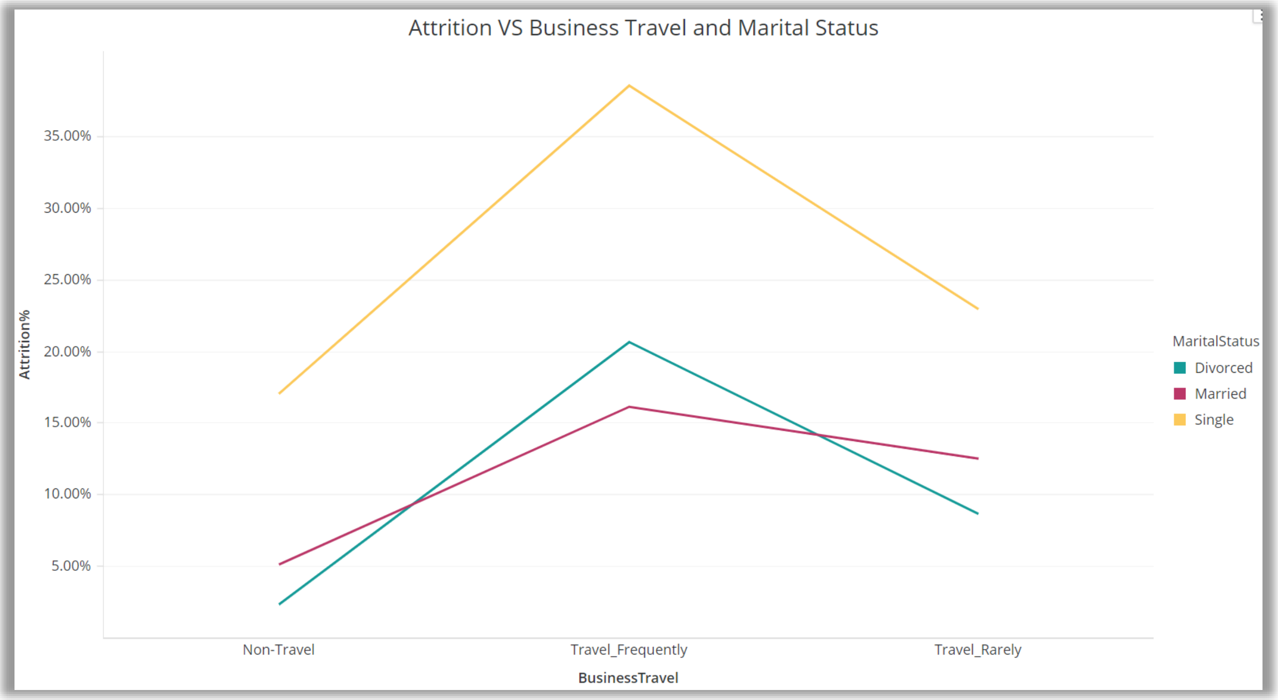
Descriptive analysis:

Visualizing the data with some charts

Figure 5 Bar chart for Attrition VS Education and Travel frequency

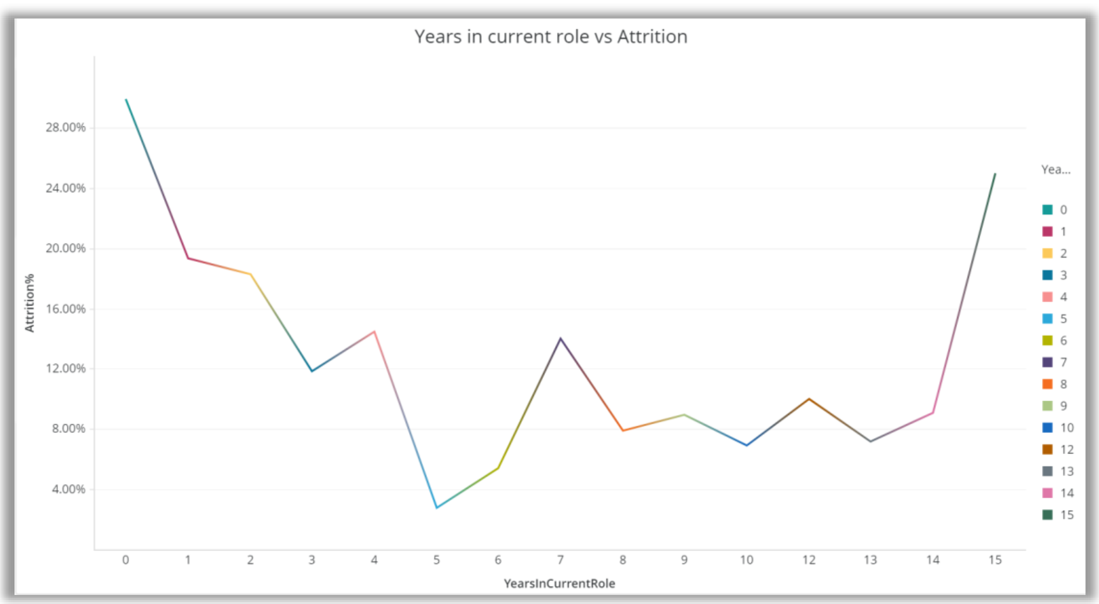


Here data shows attrition is high for any education level if the business travel is frequent.



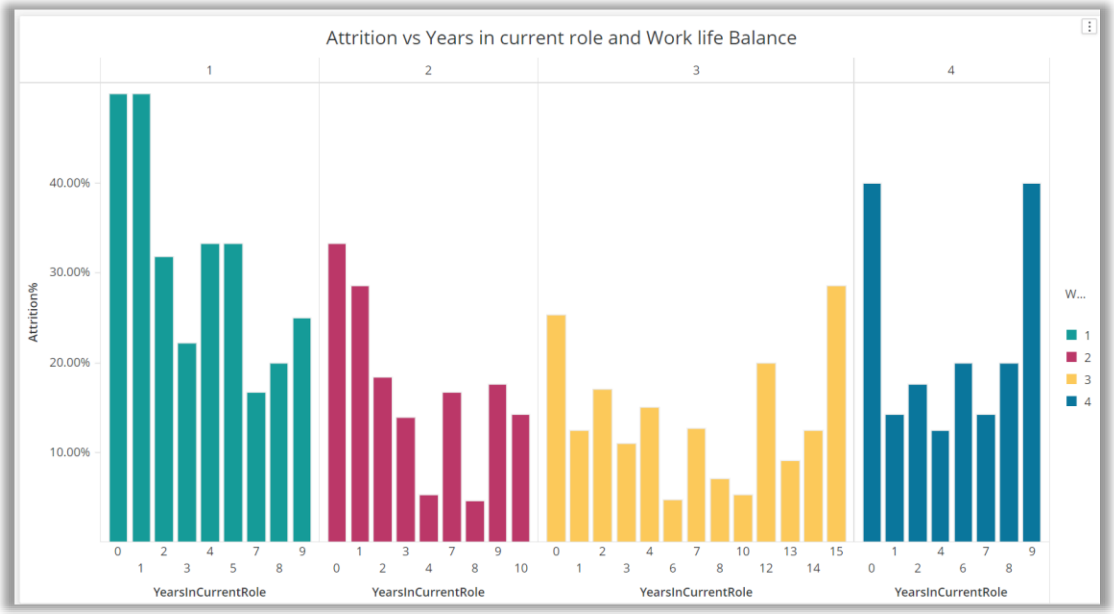
Unfortunately, organizations do not go back from pushing people who single for more business travel which can be seen from the about line chart.

Figure 6 Line chart for Attrition VS Years in current role



It also indicates that attrition is higher at 29.9% even if an employee is new and has been in the current role for less than 2 years and with employees staying longer in the same role. However, we can check if this has anything to do with the work-life balance rating.

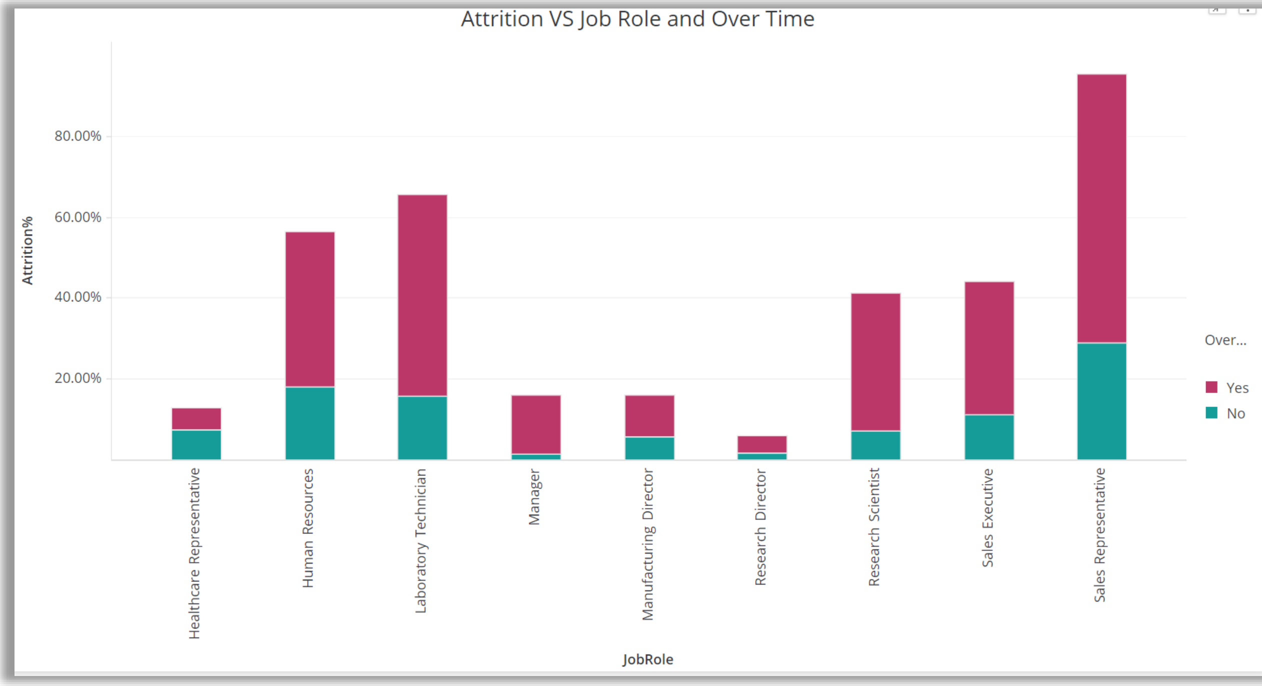
Figure 7 Bar chart for Attrition VS Years in the current role and work-life balance



We can see lower the work-life balance rating higher the attrition. For employees in the same role for a longer duration, work-life balance is not an important motivation factor.

Since we know overtime is also a factor for attrition, we can visualize which role is experiencing the most over time.

Figure 8 Stacked Bar chart for Attrition VS Jobe Role and Over Time



This explains Sales rep, Human resources and Lab techs face a higher over time as well as higher attrition rate. Sales representatives face attrition as high as 66.6%, HR 38%, lab technicians 50% and research scientist with 34%. All these visuals provide us various insights on which variables should be focused more to reduce the attrition. We also see that Attrition is high at 81% for people at lower-level jobs and they go through over time along with work-life balance at a lower rating.

Figure 9 Stacked Bar chart for Attrition VS Job level and Over Time

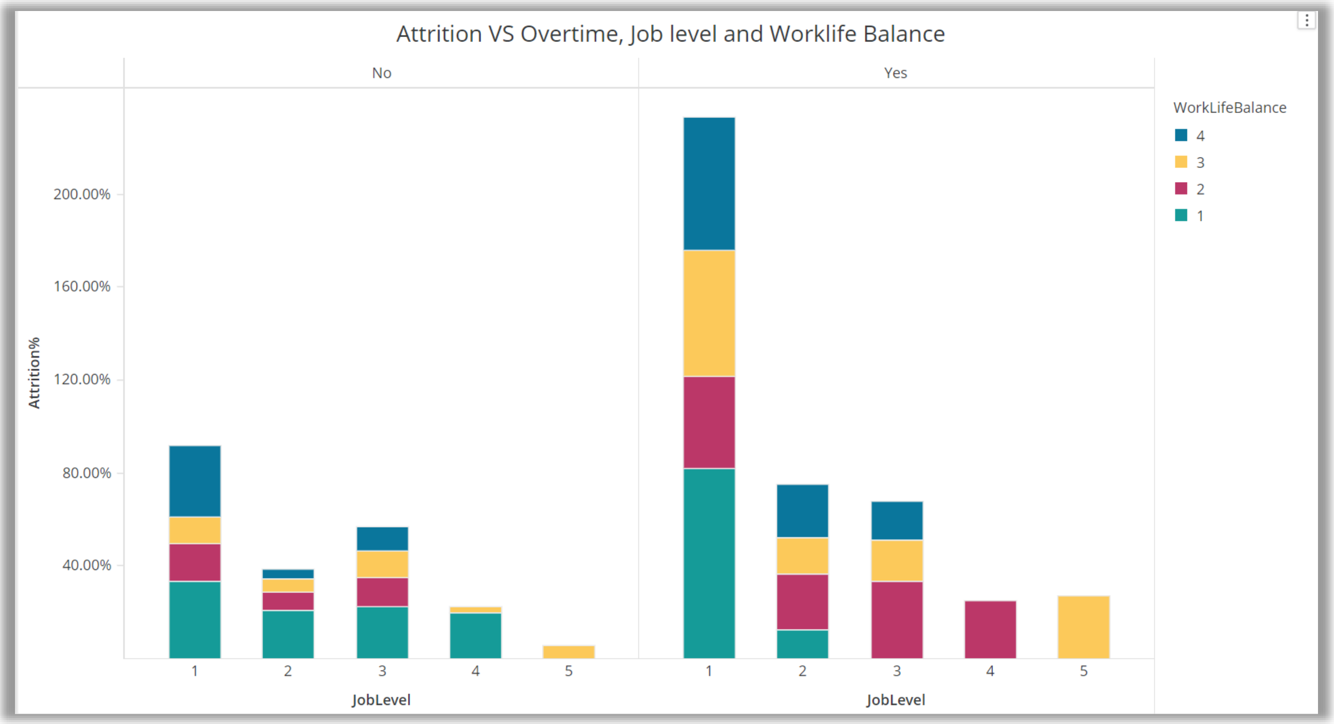
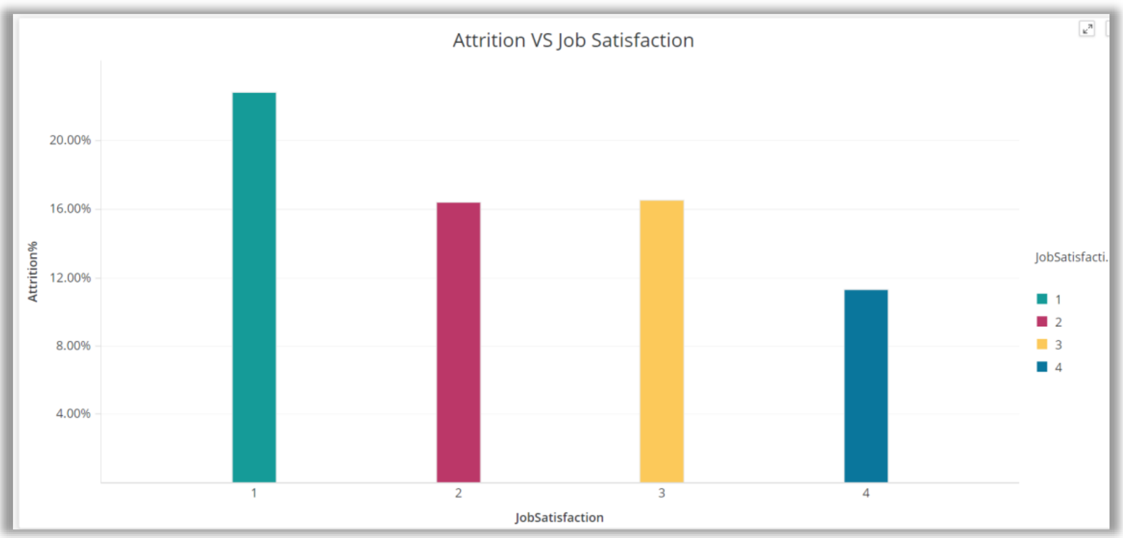
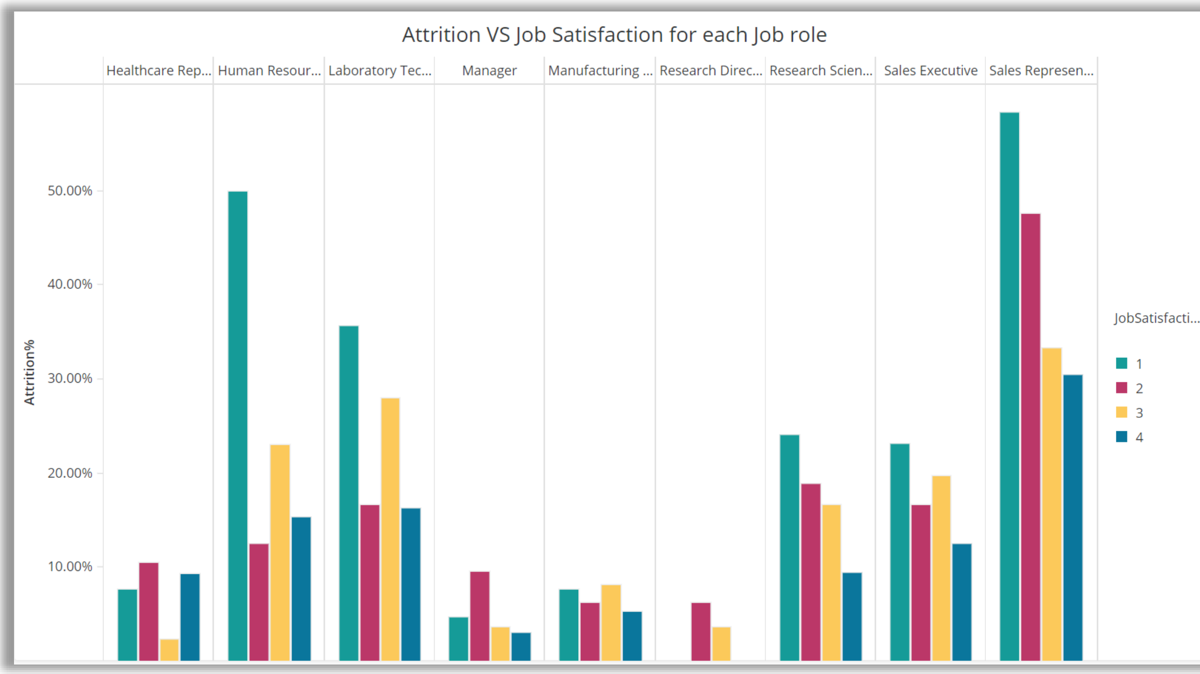


Figure 10 Stacked Bar chart for Attrition VS Job Satisfaction.



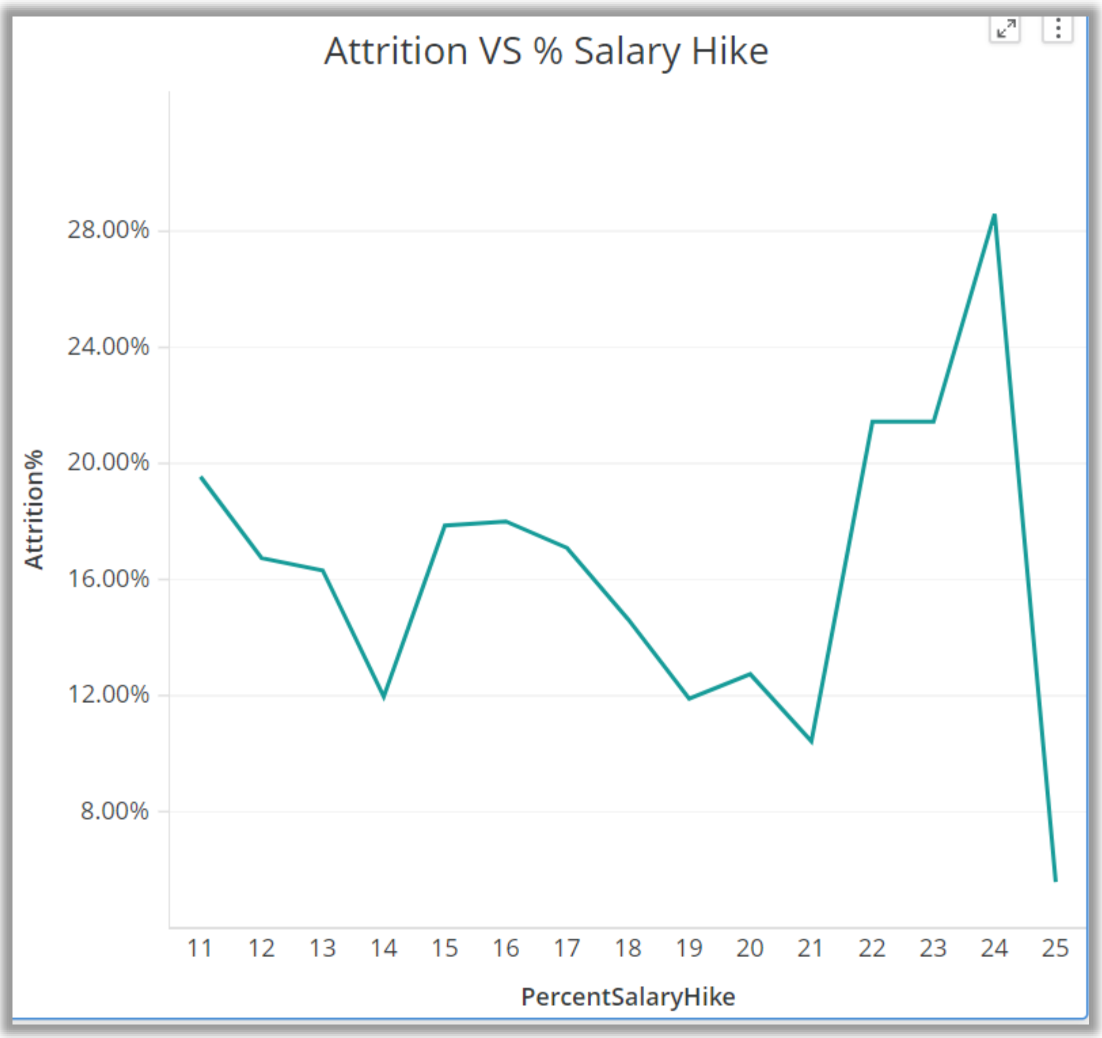
We have seen several attributes and we know Job level, job role, over time and Business travels are affecting the attrition rate.

We can also see that job satisfaction is something correlated to Attrition. The job role is correlating with Job satisfaction. We know sales and HR are one of the departments with the highest correlation. Below chart correctly indicates that Job Satisfaction is low for these two departments due to which attrition is also high.



Below chart indicates that even with high salary hike the attrition rate is high at 28%. This shows that money is not a motivating factor for the employees to stay in the company. Sometimes people wait for the hike and wait till year-end for their shift.

Figure 11 Line chart for Attrition VS Percentage Salary Hike



Predictive Modelling

Assigning a new variable for the dataset to work on Decision tree and create a new data frame using only the variables which look relevant and have a correlation with Attrition.

Figure 12 New data frame for decision tree.

```
ibmattritionDT <- read.csv("Attrition.csv", header = T, stringsAsFactors = FALSE)


attritiondf <- data.frame(cbind(ibmattritionDT$OverTime,
                                ibmattritionDT$BusinessTravel,
                                ibmattritionDT$YearsWithCurrManager,
                                ibmattritionDT$YearsSinceLastPromotion,
                                ibmattritionDT$Attrition,
                                ibmattritionDT$EnvironmentSatisfaction,
                                ibmattritionDT$JobLevel,
                                ibmattritionDT$JobRole,
                                ibmattritionDT$PercentSalaryHike,
                                ibmattritionDT$TotalWorkingYears,
                                ibmattritionDT$WorkLifeBalance,
                                ibmattritionDT$YearsInCurrentRole))

colnames(attritiondf) <- c("Overtime",
                           "Businesstravel",
                           "Yrswithmanager",
                           "Yrssincepromotion",
                          "Attrition",
                          "EnvironmentSatisfaction",
                          "Joblevel",
                          "JobRole",
                          "PercentSalaryHike",
                          "TotalWorkingYears",
                          "WorkLifeBalance",
                          "YearsInCurrentRole")
```

Using mutate function to convert these variables to factor level.

Figure 13 Mutate function

```
attritiondf <- mutate(attritiondf,
                       Overtime = factor(Overtime, levels = c("Yes", "No"), labels = c("Yes", "No")),
                      Businesstravel = factor(Businesstravel, levels=c("Non-Travel","Travel_Frequently","Travel_Rarely"),
                                            labels=c("Non-Travel","Travel_Frequently","Travel_Rarely")),
                      Yrswithmanager = as.integer(Yrswithmanager),
                      Yrssincepromotion = as.integer(Yrssincepromotion),
                      Attrition = factor(Attrition, levels = c("No","Yes"), labels = c("No","Yes")),
                      EnvironmentSatisfaction = as.integer(EnvironmentSatisfaction),
                      Joblevel = as.integer(Joblevel),
                      JobRole = factor(JobRole, levels = c("Healthcare Representative","Human Resources",
                                                          "Laboratory Technician","Manager","Manufacturing Director",
                                                          "Research Director","Research Scientist","Sales Executive",
                                                          "Sales Representative"),
                                      labels = c("Healthcare Representative","Human Resources","Laboratory Technician",
                                                 "Manager","Manufacturing Director","Research Director",
                                                 "Research Scientist","Sales Executive","Sales Representative")),
                      PercentSalaryHike = as.integer(PercentSalaryHike),
                      TotalWorkingYears = as.integer(TotalWorkingYears),
                      WorkLifeBalance = as.integer(WorkLifeBalance),
                      YearsInCurrentRole = as.integer(YearsInCurrentRole))
```

Use Glimpse function for the new data frame which will show these variables are correctly assigned as factors and integers.

Figure 14 Glimpse function

```
> glimpse(attritiondf)
Observations: 1,470
Variables: 12
$ Overtime                <fct> Yes, No, Yes, Yes, No, No, Yes, No, No, No, No, Yes, No, No, Ye...
$ Businesstravel          <fct> Travel_Rarely, Travel_Frequently, Travel_Rarely, Travel_Frequen...
$ Yrswithmanager          <int> 14, 16, 1, 1, 11, 15, 1, 1, 17, 16, 12, 17, 12, 11, 12, 17, 14,...
$ Yrssincepromotion       <int> 1, 2, 1, 10, 9, 10, 1, 1, 2, 14, 1, 1, 11, 2, 1, 15, 1, 1, 10, ...
$ Attrition               <fct> Yes, No, Yes, No, No, No, No, No, No, No, No, No, No, No, Yes, ...
$ EnvironmentSatisfaction <int> 2, 3, 4, 4, 1, 4, 3, 4, 4, 3, 1, 4, 1, 2, 3, 2, 1, 4, 1, 4, 1, ...
$ Joblevel                <int> 2, 2, 1, 1, 1, 1, 1, 1, 3, 2, 1, 2, 1, 1, 1, 3, 1, 1, 4, 1, 2, ...
$ JobRole                 <fct> Sales Executive, Research Scientist, Laboratory Technician, Res...
$ PercentSalaryHike       <int> 1, 13, 5, 1, 2, 3, 10, 12, 11, 3, 3, 2, 7, 1, 4, 1, 2, 3, 6, 1,...
$ TotalWorkingYears       <int> 39, 3, 38, 39, 37, 39, 5, 2, 3, 10, 37, 3, 36, 24, 37, 3, 38, 2...
$ WorkLifeBalance         <int> 1, 3, 3, 3, 3, 2, 2, 3, 3, 2, 3, 3, 2, 3, 3, 3, 2, 2, 3, 3, 2, ...
$ YearsInCurrentRole      <int> 14, 17, 1, 17, 12, 17, 1, 1, 17, 17, 14, 15, 12, 12, 12, 19, 12...
```

Creating a data frame to test and train for decision tree. Assigning 80 % data to the train table and 20% to test of the 1470 observations.

Figure 15 Train and test data frame

```
> attritiontrainDT <- attritiondf[1:1176,]
> attritiontestDT <- attritiondf[1177:1470,]
```

Using prop.table function to check if the randomization process is right.

```
> prop.table(table(attritiontrainDT$Attrition))

      No      Yes 
0.835034 0.164966 
```

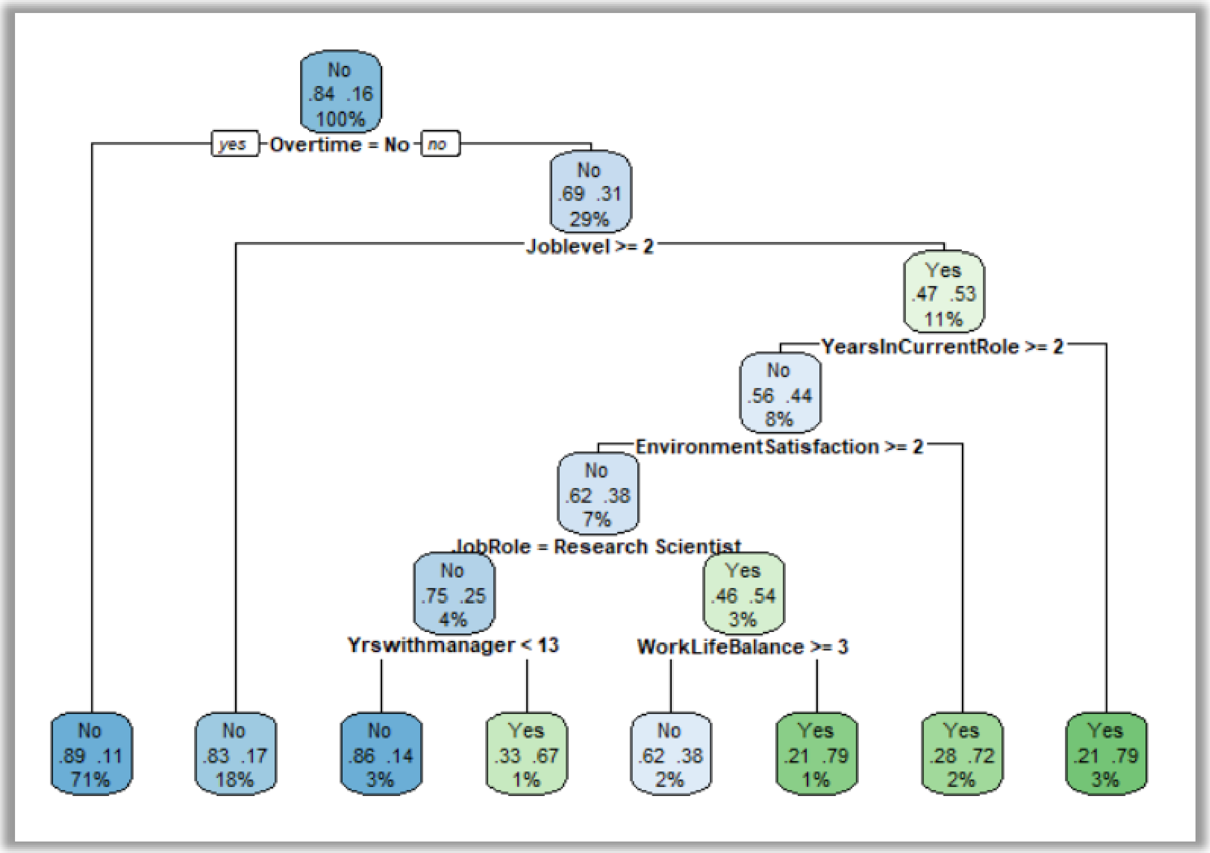
```
> prop.table(table(attritiontestDT$Attrition))

       No       Yes 
0.8537415 0.1462585 
```

Now run the function to fit the model and plot the decision tree.

Figure 16 rpart function to fit the model and plot

```
> fit <- rpart(Attrition ~., data = attritiontrainDT, method = 'class')
> rpart.plot(fit,extra = 104, cex=0.7)
```



Interpretation of Decision tree:

The first node indicates that 16% of the population in the dataset have left the organization. Out of this 16% population, we can ask the question about over time, job level, years in current role etc. If they did not do over time, then 11% of the employees left the firm and if the employees did overtime then 31% of them quit. On the other side, we can check if employees who did overtime and are at more than or equal to 2 years at the current job level then, 17% employees quit the company and 47% employees left who were doing overtime having job level less than 2 years. Most of the variable here in the tree is Overtime, Job level and job role which we could get an idea from the visuals. We can ask similar questions from the decision tree and get valuable insights. We will predict the test data frame which we created to check the accuracy of this. The model correctly predicted 16% i.e. approximately 240 employees left the organization but classified 7 who did not leave the organization which is not correct.

```
> predict_unseen <-predict(fit, attritiontestDT, type = 'class')
> temp <- table(attritiontestDT$Attrition, predict_unseen)
>
>
> temp
     predict_unseen
       No Yes
  No  244   7
  Yes  33  10
> accuracy_Test <- sum(diag(temp)) / sum(temp)
> print(paste('Accuracy for test', accuracy_Test))
[1] "Accuracy for test 0.863945578231292"
> 
```

Prescriptive Analysis:

We can make some valid recommendation based on above analysis. Variables Over time, Job role, job level, Years in the current role and business travel has a direct impact on attrition rate. Some of the Job roles and Job level have higher overtime rate which is causing the attrition. This gives us our answer for the business problem. This means management should focus more and try to alleviate the pain areas so that employees stay for a longer duration. Increase in attrition is always a loss for the organization. It impacts financially due to the amount of time and money invested on the employee is lost once he or she leaves the organization. The organization can improve the team balance so that over time can be reduced. It always happens that entry-level employees take most of the load since they are expected to prove in their initial days. However, their productivity is worn out in the first two years due to which they lose the motivation and leave the company. This can be avoided by reducing their overtime and spreading the work equally wherever possible. Traveling frequently is again another factor which makes an employee quit. A person will travel and be motivated for the initial few years. Over time their interest is reduced, and people start looking for jobs which will not have a lot of travel.

Unless required, traveling schedule should be divided equally among the team which will help to retain the employees who wish to take some break from their business travel. Adhikari (2009) asserts that it is not only important to implement strategies which will help in bringing down the attrition rates but also to identify the right candidates during the recruitment process so that they can prove as asset to the company. Culture shock is something which often leads to attrition by employees taking too much pressure from their new expectations. He has rightly pointed out that employees should be given the right amount of balance between flexible work time and work from home without effecting the productivity. The article also states that there should be equal weightage given to spoiling their employees and asking for quality of maximum amount of work.

Conclusion

Pramod, Raman & Bhattacharya (2019) asserts that predictive analysis to predict real-time behavior can give useful insights to the organization. Organizations or management's actions can influence an employee's attrition. Job postings, hiring new employees, training them again are some of the lost opportunity and cost. Not to forget the talent which is going to the competitor. As data is proving to be more and more important every day and even for human resource data can provide useful insights which can help them to retain their talent from going to their competitor. This set of data gave us many factors contributing employees to quit, which most of the organization already know. As per the author it is very difficult to fully understand the exact reason why employee quit the organization. Data from forms and exit interviews can give only a limited amount of understanding. Most of the time employees do not disclose the reason behind their switching to a different organization. The reason can also be due to layoff or health reasons which cannot be controlled by the company. As per the article "Predict employee attrition by using predictive analytics", research which was done on an information technology organization revealed that precision in the role and projects which play important role in increasing employees' loyalty towards the company is usually the key factor. Khera & Divya (2019) explained in their predictive analysis that attrition is influenced by various factors like personal, work related, organizational policies etc. these factors will help us to predict their turnover in the organization. Companies most of the time conduct survey in and around their performance appraisal cycle to understand employees' attitude towards the company with that predict the percentage of attrition for that year. Survey results can be analyzed with cluster analysis as well. Predictive analytics can help us to understand the data and give us answers which will help to retain the talent in future at a similar behavior by a different group. In this study with the help of decision tree, we learned the variables which matters the most to employees and due to which they are motivated to leave. The different variables and number of observations allowed to create test and train model which gives us better result for predicting the attrition.

Reference

Adhikari, A. (2009). Factors Affecting Employee Attrition: A Multiple Regression Approach. ICFAI Journal of Management Research, 8(5), 38–43. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=bth&AN=39231799&site=eds-live&scope=site

Allen, D (n.d.) Retaining Talent: A Guide to Analyzing and Managing Employee Turnover [PDF]. Retrieved from https://www.shrm.org/hr-today/trends-and-forecasting/special-reports-and-expert-views/Documents/Retaining-Talent.pdf

Anderson, B (2018, November 14). It's Time To Focus On Your Employee Retention Strategies. Retrieved from https://www.forbes.com/sites/forbescommunicationscouncil/2018/11/14/its-time-to-focus-on-your-employee-retention-strategies/#19da2ff36ba4

Conerly, B. (2018, August 12). Companies Need To Know The Dollar Cost Of Employee Turnover. Retrieved from https://www.forbes.com/sites/billconerly/2018/08/12/companies-need-to-know-the-dollar-cost-of-employee-turnover/#7d44449d590a

Dobhal, R & Nigam, A. (2018, February). Employee Attrition and Employee Satisfaction. A Study of HR Performance Appraisal. IOSR Journal of Business and Management (IOSR-JBM) e-ISSN: 2278-487X, p-ISSN: 2319-7668. Volume 20, Issue 2. Ver. IV (February. 2018), PP 01-27 www.iosrjournals.org

Frye, Alex; Boomhower, Christopher; Smith, Michael; Vitovsky, Lindsay; and Fabricant, Stacey (2018) "Employee Attrition: What Makes an Employee Quit?," SMU Data Science Review: Vol. 1 : No. 1 , Article 9.

Khera, S. N., & Divya. (2019). Predictive Modelling of Employee Turnover in Indian IT Industry Using Machine Learning Techniques. Vision (09722629), 23(1), 12–21. https://doi.org/10.1177/0972262918821221

Kaggle (n.d.) IBM HR Analytics Employee Attrition & Performance. Retrieved from https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

Plonski, P (2017, April 27). Human Resources Analytics: Predict Employee Attrition. Retrieved from https://medium.com/@MLJARofficial/human-resources-analytics-predict-employee-attrition-5ddc3ed781c

Raman, R., Bhattacharya, S., & Pramod, D. (2019). Predict employee attrition by using predictive analytics. Benchmarking: An International Journal, 26(1), 2. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edb&AN=134654379&site=eds-live&scope=site

Sen Gupta, S (2018). Employee Attrition and Retention: Exploring the Dimensions in the Urban Centric BPO Industry [PDF]. Retrieved from https://pdfs.semanticscholar.org/6913/381a3124a5221c927c0a50719d4e69615c6e.pdf