

# Business Analytics with R: Employee Churn Project Report

## Project Description:

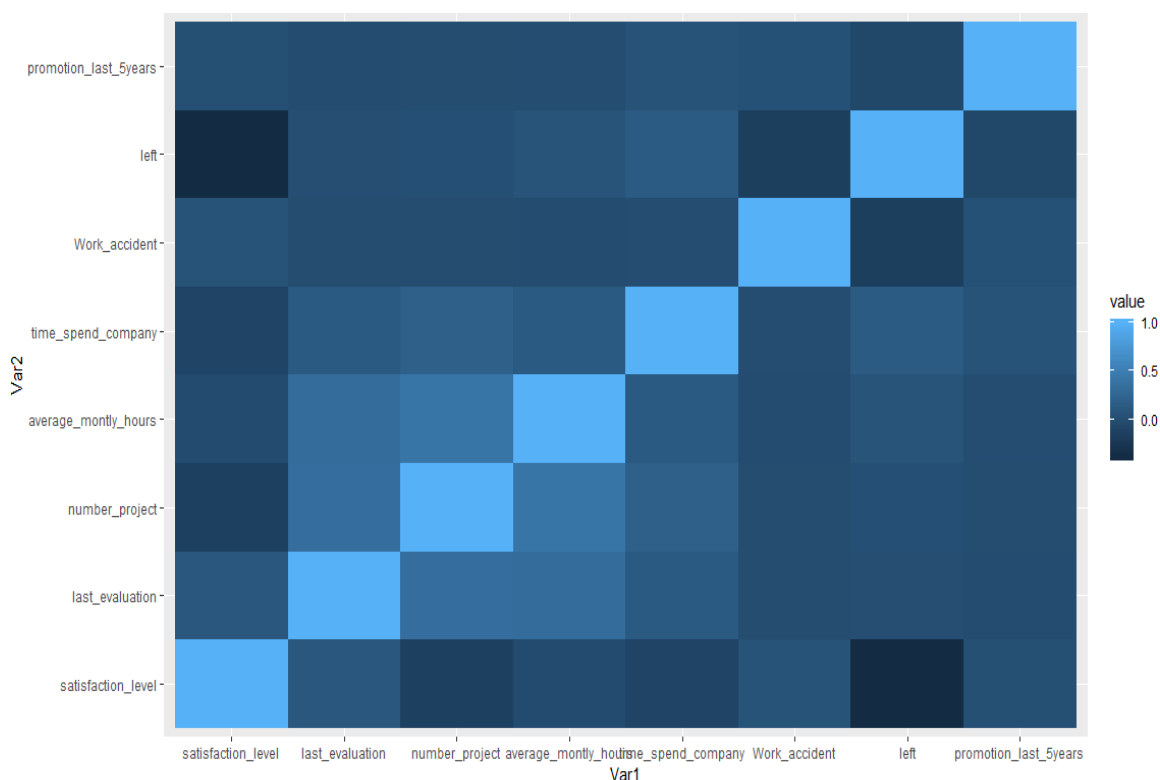
The dataset is an employee churn dataset that was found on Kaggle. The dataset has 14,999 rows and 10 columns. Each row represents a different employee, and the columns represent the satisfaction level, last evaluation, number of projects, average monthly hours, time spent with company, work accidents, left (left the company), promotion within last 5 years, sales, and salary.

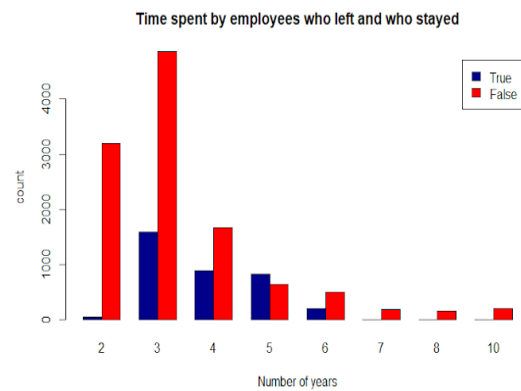
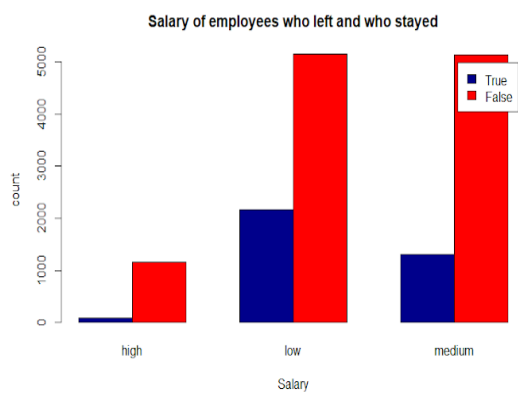
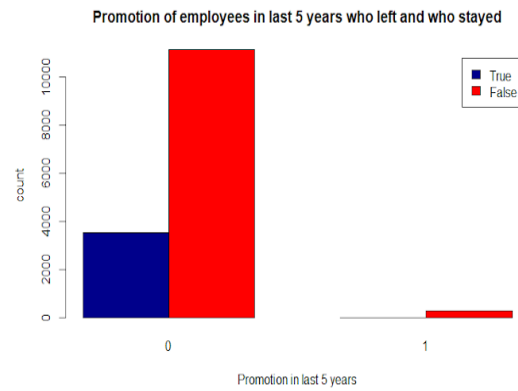
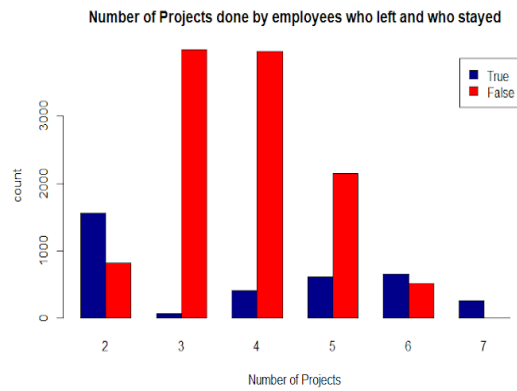
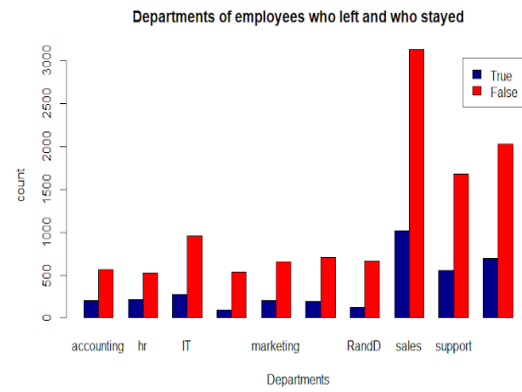
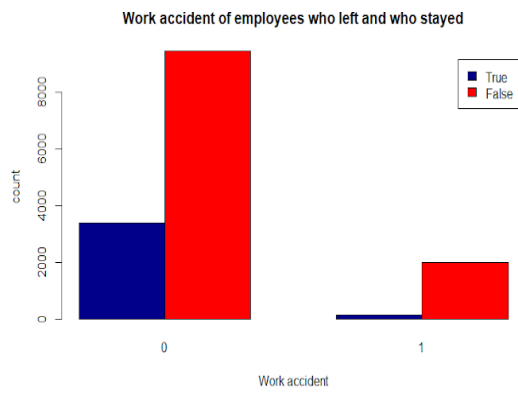
The objective of this project is to analyze the human resources dataset and find the cause of employee turnover within the dataset, as well as predict if an employee will leave the organization or not based on the other attributes about the employee. In terms of the dataset that we will be using, our group will be predicting if an employee has left the company (left = 1) using all the other columns within the dataset.

## Data Exploration:

### Heatmap :

- 76% of the employees stayed and 24% employees left the company
- From the heatmap, there is a positive(+) correlation between number\_project, averageMonthlyHours, and last\_evaluation.
- This leads to the obvious conclusion that the employees who spent more hours and did more projects were evaluated highly.
- For the negative(-) relationships, left and satisfaction are highly correlated followed by promotion in last 5 years.
- We can say that people tend to leave a company more when they are less satisfied and not promoted for a about 5 years.





**Interpretations :**

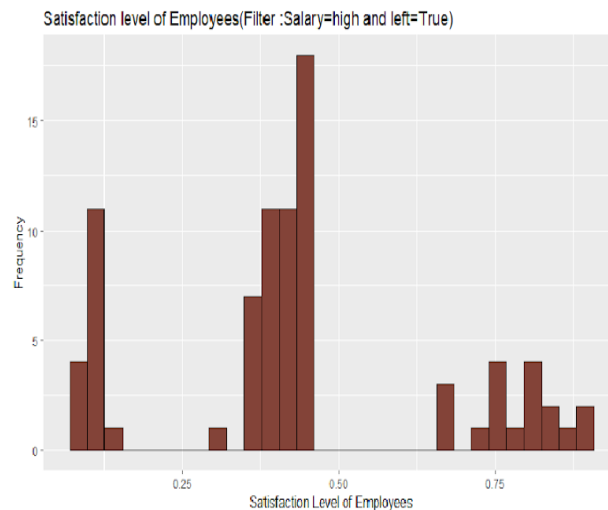
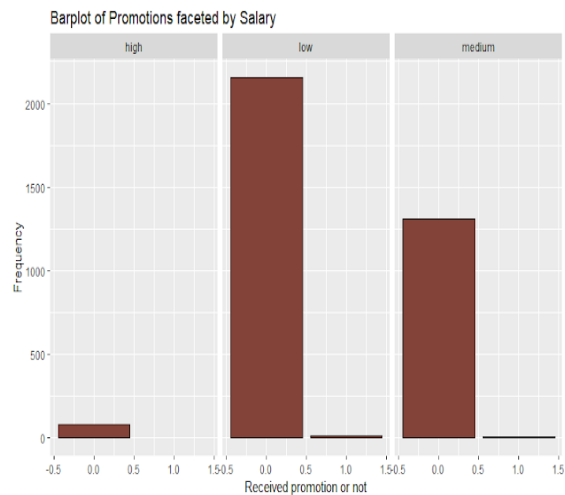
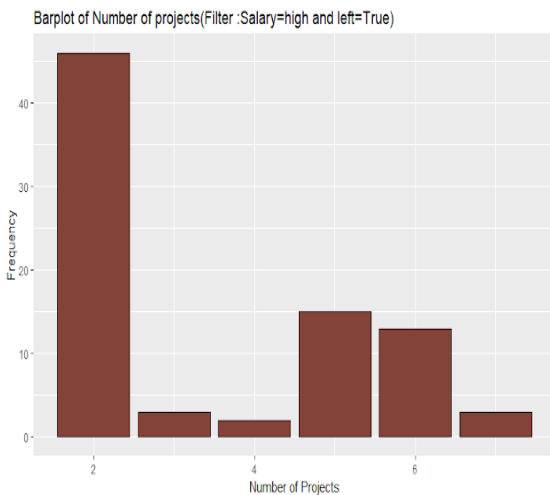
- Employees left even with 0 work accident, which implies that work accident is not an important reason which affects the churn.
- Although sales department has highest number of employee, the highest churn rate is of the HR department (29%) followed by accounting and technical. This could mean that because of low number of employees in the department, the workload is more and hence the high churn.
- The number of employees with 2 projects who left is more than who stayed. Maybe because they weren't much satisfied with their work and had low salary.
- Employees with 6 and 7 projects may have left because of extra workload.
- The employees with 3-5 years of experience are leaving more because of no promotions in last 5 years
- Employees with more than 6 years' experience are not leaving because they are promoted and have medium or high salary.
- Those who were promoted did not leave. Many of those who were not promoted left because of low salary and low satisfaction

## **Findings :**

### **1) Few Employees left the company in spite of getting highly paid**

Following could be the possible reasons:

- The plots show the count of employees who have left the company
- Even with high salary, many employees had only 2 projects and so they left. These employees do not seem to enjoy the work and expected more number of projects
- On the other hand, people with more number of project seem to have extra workload and low satisfaction
- They were not promoted in last 5 years
- Average satisfaction level is low

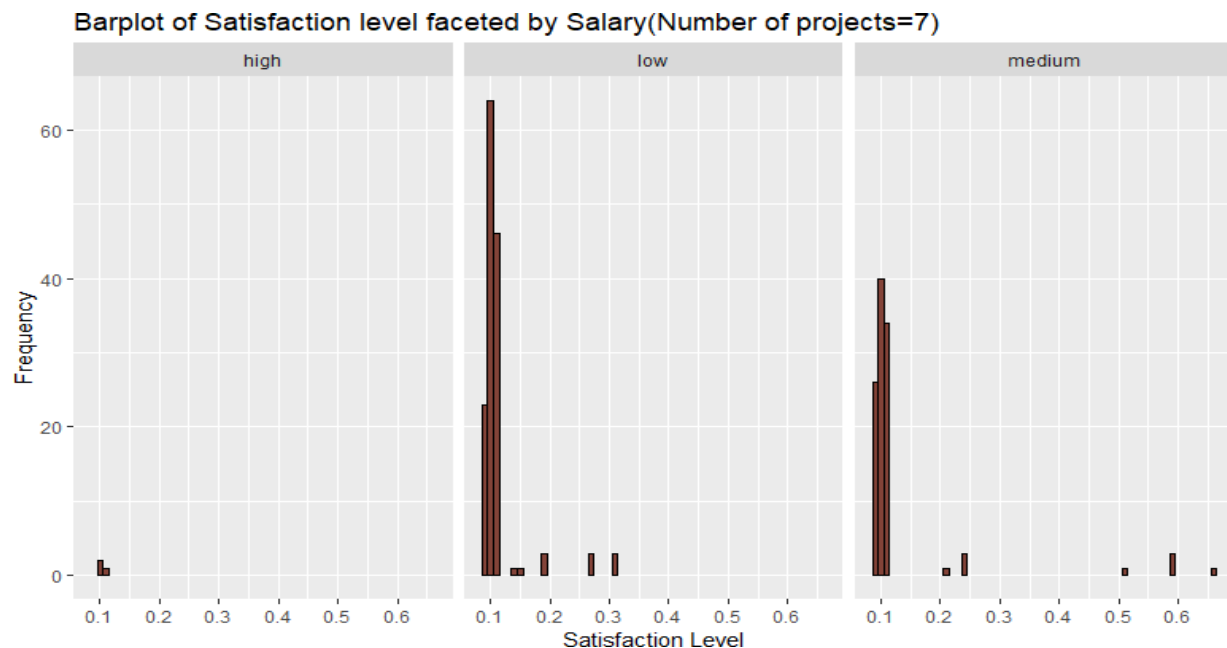


## 2) All employees who completed 7 projects left the company

Following are the possible reasons:

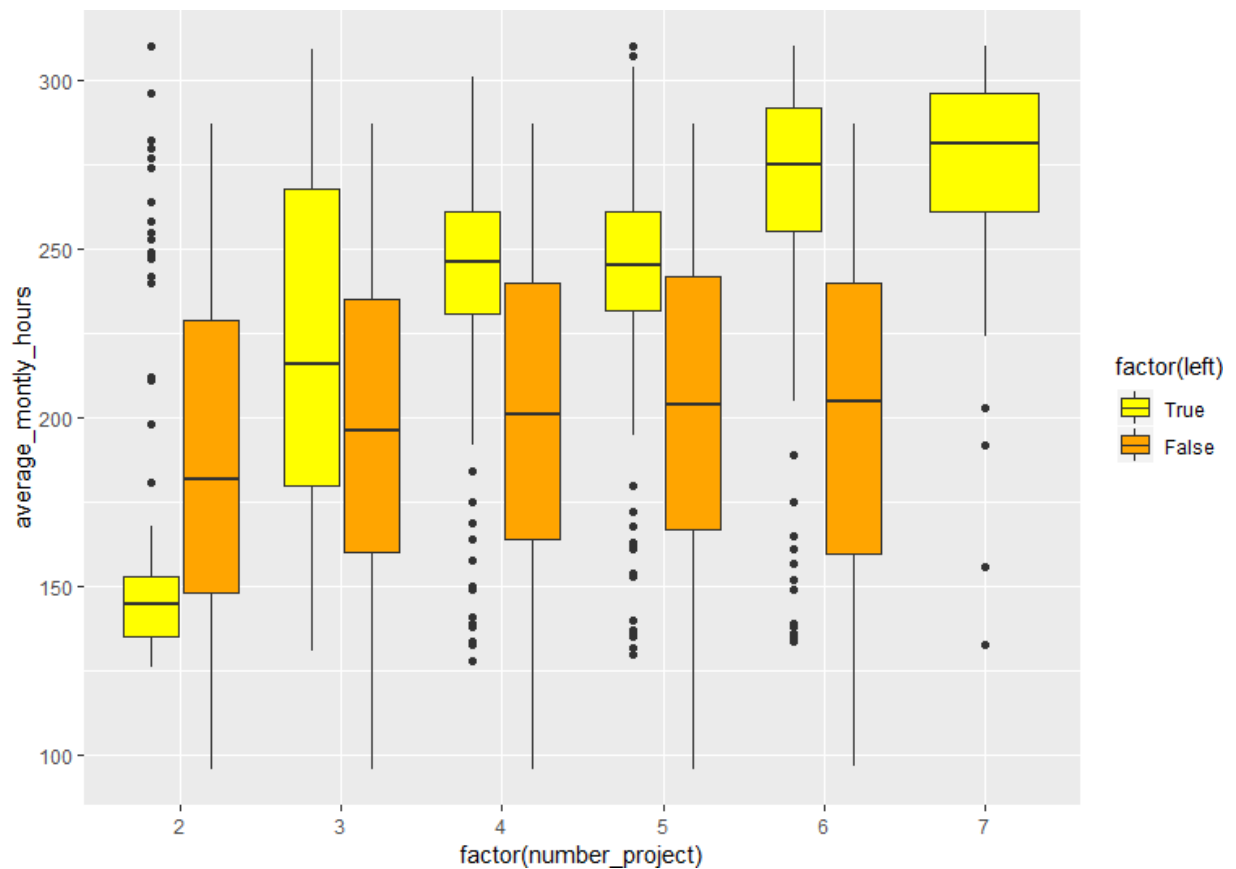
- no promotion was given even with good employee evaluation
- majority had low or medium salary
- average monthly hours spent is high
- average satisfaction level is 0.12

```
> project7 <- filter(hrm, number_project == 7)
> result <- aggregate(x = project7[,c(-7,-9,-10)],
+                     by = list(project7$left),
+                     FUN = mean)
> result
  Group.1 satisfaction_level last_evaluation number_project average_monthly_hours time_spend_company
1   True      0.1187109      0.8596094              7          276.0781             4.113281
  Work_accident promotion_last_5years
1    0.015625              0
>
>
```



3)

- As project count increased, so did average monthly hours
- Employees who did not have a turnover had consistent average Monthly Hours, despite the increase in projects
- In contrast, employees who did have a turnover had an increase in averageMonthlyHours with the increase in projects
- The average working hours of the employee is around 200. Majority of the employees who left the company had working hours of 250

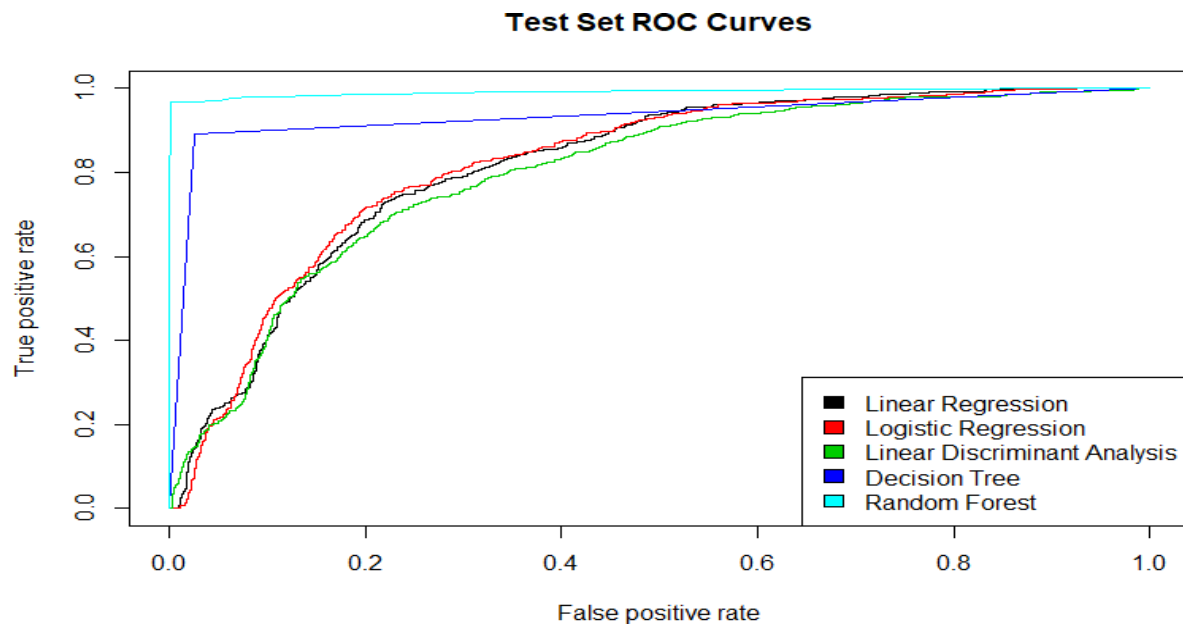


## Model Selection :

On comparing the performances of each of the models by using an ROC curve, we can conclude the Random forest and Decision tree models are best suited for predicting the turnover of employees.

A better selection between the two would be to use Decision tree model as it allows for easier and clearer understanding of classifications and a comprehensive view of the attributing factors of employee churn.

The decision to select the model often falls on the data scientist factoring in computation power and size of dataset and level of comprehension that is required.



### Model accuracy Table

Index	Model name	Accuracy (percentage)	Area under the curve
1	Linear Regression	76.892 ~ 77%	0.823
2	Logistic Regression	77.959 ~ 78%	0.828
3	Linear Discriminant Analysis	77.359 ~ 77%	0.8101
4	Decision Tree	96.47 ~ 97%	0.9475
5	Random Forest	99.3 ~ 99%	0.9923



## **Recommendations:**

- From our analysis we also came up with some recommendations for companies who want to retain their employees.
- We saw from the dataset that a very high number of projects was related to the employee leaving the company. Also, a very low number of projects was related to the employee leaving.
- Employee satisfaction also seemed to play a major role in the employee's decision to leave. Therefore, a company should try to measure various attributes related to the employee's safety, satisfaction, workload, etc. to try and understand if an employee will leave and utilize that information to increase employee satisfaction levels or decrease workload for an unsatisfied employee.
- Many companies overlook attributes related to job satisfaction of their employees, and ultimately lose talent within the organization and it can prove to be costly.
- Promoting an inclusive and comfortable work environment has proven to help retain employees
- In addition to our algorithms, we can conclude that Maternity leave, work places with childcare can help decrease churn rate.
- Promoting a healthy work-life balance directly affects the satisfaction levels of employees.
- Acknowledging and recognizing excellence with rewards makes a good motivating factor for employees to stay on
- Long hours at the workplace due to understaffing can result in unhappy employees. Therefore the companies must ensure that they are adequately staffed to prevent overworking their employees