A Fresh New Perspective on Employee Satisfaction and Retention

FEBRUARY 15, 2019

IIM-A EPABA BATCH2 - GROUP 17

UNDER THE GUIDANCE OF PROF. ANKUR SINHA

AUTHORED BY:

VIKAS GOEL

NILIMA GHOSH

SRIKANTH SHETTY

Table of Contents

Introduction 3

Problem Statement 3

Data 4

Definitions of each variable 4

Methodology 5

1. Deep Dive and Understanding of data using Exploratory Analysis & Data Visualization 5
2. Data Pre-processing 5
3. Data split (Train/Test) 5
4. Created predictive models: Binary Classification problem 5
5. Comparing Models and choosing the best model 6
6. Evaluate Feature importance to know why employees are leaving 6
7. Create Model for satisfaction Level 6
8. Conclusions & Recommendations 6

Results 6

Result Discussions 6

Cost Analysis 6

Choosing Best Model 7

Who is leaving? - Decision Tree: 7

Why are they leaving? – Feature importance: PCA, Random forest 7

Conclusions 8

Why Are Employees Leaving? 8

Identify High Performing and Experienced Employees 8

The scope of Further work 9

Other Factors to be considered for calculating *the Satisfaction* of an Employee 9

Other Factors to be considered for calculating the *Real* Cost of losing an employee: 9

Risk Mitigation Steps 9

Recommendations 9

Acknowledgments 10

References 11

Appendix 12

EDA Plots 12

R Code and Results 16

Introduction

When employees leave the company, there is a multifold impact. There is quantifiable economic loss since it costs more to lose employees. According to the data drawn from various research papers, it costs additional 30-50% of their wages when the employee leaves. These costs reflect the loss of productivity after their departure, replacement cost, and the reduced productivity while the new employee gets up to speed. With multiple employees leaving every year, there is not only big dent in the budget, but it also is detrimental to the moral of the current employees working in the company. Hence, it is important for organizations to find why their first-class employees are leaving prematurely and to predict who could be leaving the organization next. This will help them to create policies to improve Employee Retention.

Hence, the primary focus of the project is to understand why valuable employees are leaving the organization, understand how satisfaction level plays a role in an attrition, predict the employees who might be leaving the next and find a way to retain them.

Problem Statement

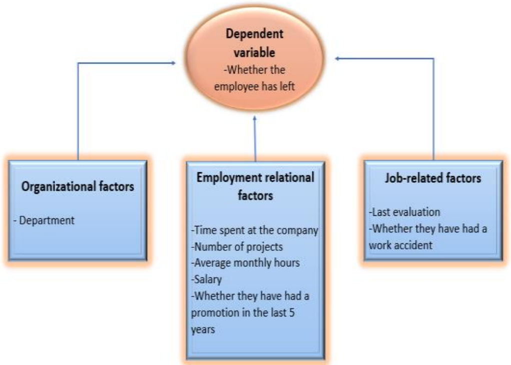
- Employees are the most significant assets of an organization; customers come second. Being able to hire and retain the right talent enhances a company's reputation.
- Moreover, employee turnover (attrition) is a significant cost to an organization, and predicting turnover is at the forefront of the needs of Human Resources (HR) in many organizations.
- There are a few cases when employees quit their organizations without giving any signal. Except for the human factors, it is possible to notice indicators or signs when employees are contemplating their exit.
- Identifying high performing employees with a plan to retain them is as important as avoiding overspend on those employees with a low probability of staying with the organization.

Data

*****Disclaimer:** The basic principle governing data privacy describes that "***Organizations cannot collect any data about an individual without the employee's specific consent about the information being collected and awareness on the purpose for which it will be used***." This principle limits the availability of employee data. We have sourced this handcrafted employee data from https://www.kaggle.com/lnvardanyan/hr-analytics

Data features – The dataset has 14999 rows and 10 attributes. The variable "left" is the response indicating whether an employee had left the company or not.

Definitions of each variable:


Dependent variable
-Whether the employee has left
Organizational factors
- Department
Employment relational factors
-Time spent at the company
-Number of projects
-Average monthly hours
-Salary
-Whether they have had a promotion in the last 5 years
Job-related factors
-Last evaluation
-Whether they have had a work accident

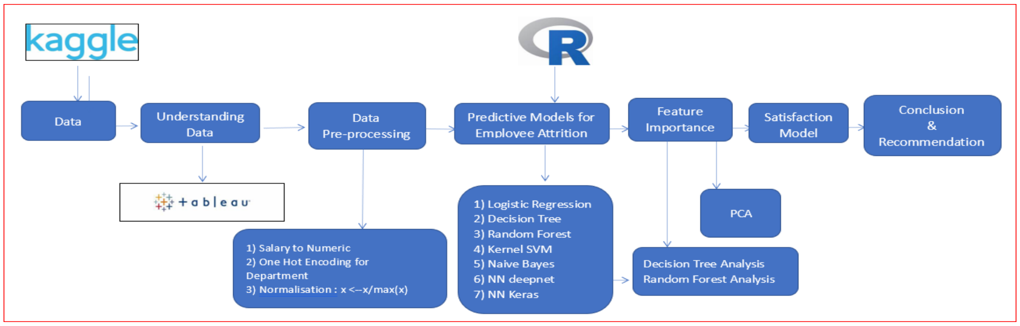
- ***satisfaction_level****:* indicates the level of the satisfaction for each employee collected from survey, 1 indicates the highest level of satisfaction
- ***last_evaluation***: the results of performance for each employee, 1 indicates the highest appraisal scores
- ***number_project:*** the total number of projects an employee had been working on
- ***average_monthly_hours***: the average monthly working hours for each employee
- ***time_spend_company***: average number of hours an employee stays in company for each working day
- ***Work_accident***: binary, whether an employee had encountered accidents in workplace
- ***left:*** the response variable, binary, indicates if the employee had left the company
- ***promotion_last_5years***: binary, whether the employee had been promoted in the past five years
- ***Department:*** indicating the function/department the employees works for
- ***salary:*** range of salary level, divided into groups of: low medium and high

	Attributes	Value range	Type of data	Remarks	Mean	Median
Continuous Numerical	satisfaction_level	0 to 1	Quantitative (Numerical) - Continuous	More of a measure	0.613	0.64
	last_evaluation	0 to 1	Quantitative (Numerical) - Continuous	More of a measure	0.716	0.72
Discrete Numerical	number_project	Positive integer - 2,3,4,5,6,7	Quantitative (Numerical) - Discrete	Countable	3.803	4
	average_montly_hours	Positive integer - 96-310	Quantitative (Numerical) - Discrete	Tend to round off to the nearest integer in hours. Will group them together.	201.1	200
	time_spend_company	Positive integer - 2,3,4,5,6,7,8,10	Quantitative (Numerical) - Discrete	Tend to round off to the nearest integer in hours. Will group them together.	3.498	3
Binary (0,1)	Work_accident	0 = No Accident 1 = Accident	Qualitative (Categorical)	Nominal. Binary values.	0.145	0
	left*	0 = Still with company 1 = Left	Qualitative (Categorical)	Nominal. Binary values.	0.238	0
	promotion_last_5years	0 = Not Promoted 1 = Promoted	Qualitative (Categorical)	Nominal. Binary values.	0.0213	0
Categorical	Department	One of the 10 values	Qualitative (Categorical)	Nominal--10 departments: accounting, HR, IT, management, marketing, product_mng, RandD, sales, support, technical	-NA	-NA
	Salary	One of the 3 values	Qualitative (Categorical)	Ordinal – Low, Medium , High	-NA	-NA

*** Output Variable**

For data exploration Tableau was used to visualize the dataset: Explained in Appendix *EDA Plots*

Methodology



We performed below steps in the execution of the project

1. Deep Dive and Understanding of data using Exploratory Analysis & Data Visualization

As a first step, using EDA, we identified employee signals to understand the likelihood of a person to stay/exit. We used Tableau and ggplot (R Library). EDA is explained in Appendix ***EDA Plots***

2. Data Pre-processing

- As a first step, the data available was explored for any missing/null values and outliers. Fortunately, our dataset was free from missing values and outliers.
- Two categorical attributes (Department and Salary) have been converted to numerical. Department converted to individual departments with 'One Hot Encoding.' 'Salary' Low, Medium & High converted to 1,2,3 respectively.
- The data range is explained as follows; Number of Projects [2 to 7], Time Spend company [2 to 6], Average Monthly Hours [96 to 304], Remaining [0 to 1]. Since there was no negative data, we decided to normalize data by dividing each maximum of that attribute to get each data in [0,1] range.

3. Data split (Train/Test)

Split the dataset into Train and Test into 70:30 ratio with random seed (123). Models are trained on train dataset and validation with test dataset.

4. Created predictive models: Binary Classification problem

To predict whether a given employee will leave the organization or not, created below models.

Model	Library:function()
Logistic Regression	glm()
Decision Tree	rpart::rpart()
Random Forest	randomForest::randomForest() # 50 trees
Kernel SVM	e1071::svm()
Naïve Bayes	e1071::naiveBayes()
Neural Network : deepnet	deepnet::nn.train() # 3 hidden layers of 50, 30, 10 neurons respectively
Neural Network : Keras	keras::keras_model_sequential(),keras::compile # 4 hidden layers 100, 50, 25,10 Refer Code for special instructions to install keras library.

Note: All models are created and tested in R 3.5.2 and RStudio 1.1.463 on Windows 10.

5. Comparing Models and choosing the best model

After getting results (confusion matrix) from all models, we compared the models using metrics like Accuracy, Area Under Curve of ROC curve, Precision, Recall and cost to the company.

6. Evaluate Feature importance to know why employees are leaving

Principal Components Analysis (PCA) was performed to get the feature importance. Also got important features from Decision tree and random forest.

7. Create a Model for satisfaction Level

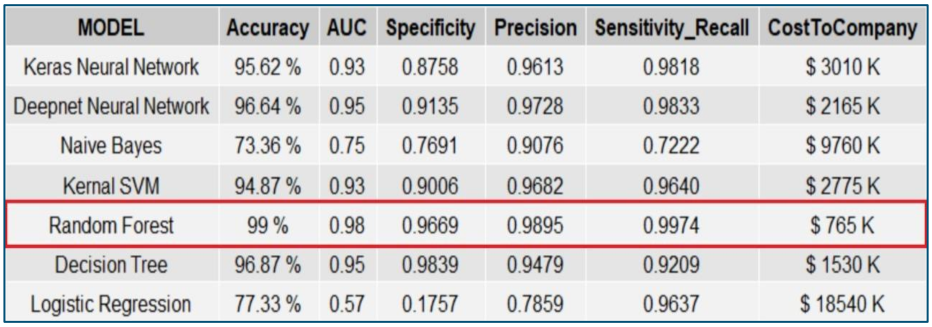
Tried creating models for satisfaction level but not getting good results. We need more attributes to have a fair model for satisfaction level. This can be future work to capture more data and do satisfaction level modeling.

8. Conclusions & Recommendations

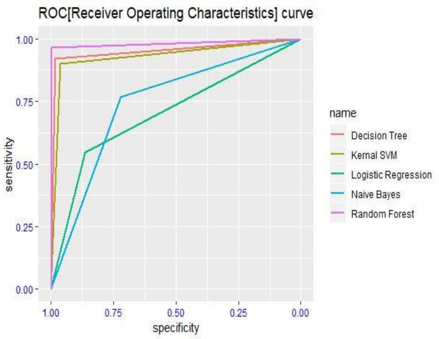
Finally used the insights to help HR understand the workplace scenario and employees behavioral attribute.

Results

Below is the result comparison of all the models which clearly indicates that we got the best results from "**Random Forest**" model. Below ROC curve, indicates the performance of each model.



MODEL	Accuracy	AUC	Specificity	Precision	Sensitivity_Recall	CostToCompany
Keras Neural Network	95.62 %	0.93	0.8758	0.9613	0.9818	$ 3010 K
Deepnet Neural Network	96.64 %	0.95	0.9135	0.9728	0.9833	$ 2165 K
Naive Bayes	73.36 %	0.75	0.7691	0.9076	0.7222	$ 9760 K
Kernal SVM	94.87 %	0.93	0.9006	0.9682	0.9640	$ 2775 K
Random Forest	99 %	0.98	0.9669	0.9895	0.9974	$ 765 K
Decision Tree	96.87 %	0.95	0.9839	0.9479	0.9209	$ 1530 K
Logistic Regression	77.33 %	0.57	0.1757	0.7859	0.9637	$ 18540 K



Note: Refer RMarkdown output for more details like confusion Matrix of each model.

Result Discussions

Cost Analysis

We considered $5k as the cost of retaining an employee and 4 times of this cost for back-filling of an employee who left the organization. With this, we calculated Extra cost to the company due to wrong predictions of the model with the formula: ***CostToCompany = ($5K * FalsePositive) + (4*$5K * FalseNegative)***

With this, we got that **Random Forest** gives minimum Cost (refer table in the above section "Result")

Choosing Best Model

On studying the model comparison table in section “Result”, we have observed that

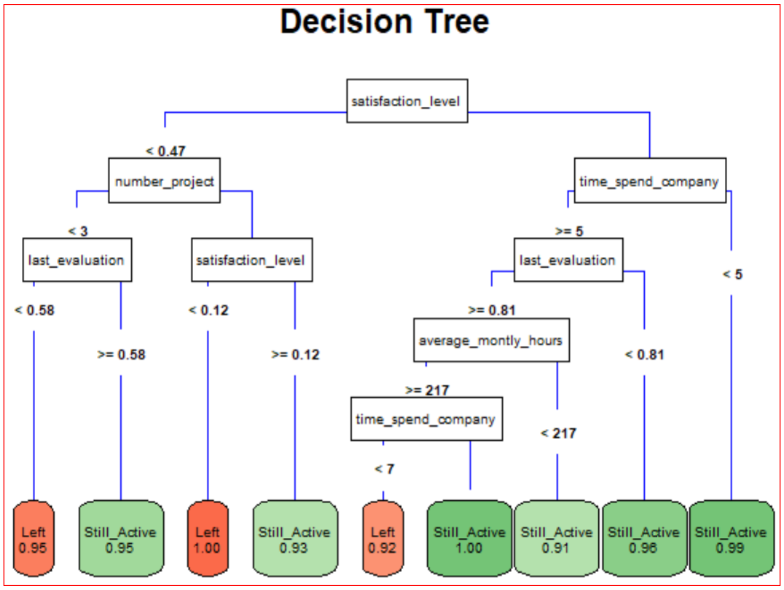
- Random Forest has the best accuracy
- Random forest has the best Sensitivity
- Random forest misclassification costs least to the company

Based on the above observations, we have chosen **Random Forest as our Final Model**.

Who is leaving? - Decision Tree:

The primary objective is to uncover employee turnover and prediction of turnover. Visual tree model (Decision Tree) explains why employees are leaving.

Satisfaction level is the most significant feature.



- **Very Low Satisfied employees**

  (Satisfaction less than 0.12): Everybody is leaving after completing 3 projects.

- **Medium satisfied employees**

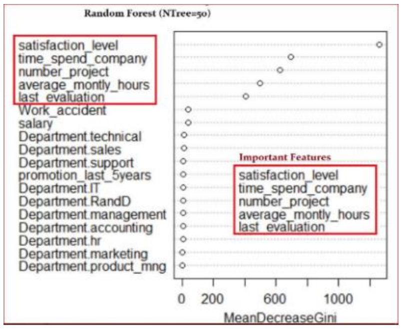
  (Satisfaction greater then 0.12 and less than 0.47): Employees will most likely leave if they have worked on less than 3 projects got less than .58 ratings.

- **Very High Satisfied employees**

  (Satisfaction> .81): They will most likely leave if they spent 5 or 6 years in the company and are spending a lot of time in the company (spending greater than 217hrs in a month). Typically, 160-180 hrs in the month is considered as balanced hrs.

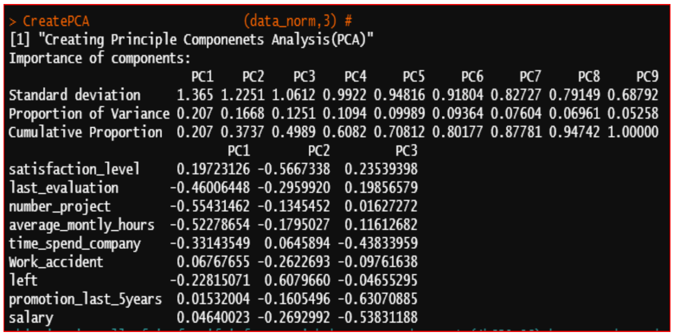
Why are they leaving? – Feature importance: PCA, Random forest

Below Mean Decrease Gini plot from Random Forest Model gives important features



1. Satisfaction Level
2. Time spent in the company
3. Number of projects completed
4. Numbers of hours spending in the company
5. Performance Ratings

PCA: In order to get more insights from data, we performed PCA on the complete dataset (excluding department). We studied the loading of the first three components which explains around 50% of data. Below is the analysis of 3 principal components.



```
> CreatePCA                    (data_norm,3) #
[1] "Creating Principle Componenets Analysis(PCA)"
Importance of components:
                         PC1    PC2    PC3    PC4     PC5     PC6     PC7     PC8     PC9
Standard deviation     1.365 1.2251 1.0612 0.9922 0.94816 0.91804 0.82727 0.79149 0.68792
Proportion of Variance 0.207 0.1668 0.1251 0.1094 0.09989 0.09364 0.07604 0.06961 0.05258
Cumulative Proportion  0.207 0.3737 0.4989 0.6082 0.70812 0.80177 0.87781 0.94742 1.00000
                               PC1        PC2         PC3
satisfaction_level     0.19723126 -0.5667338  0.23539398
last_evaluation       -0.46006448 -0.2959920  0.19856579
number_project        -0.55431462 -0.1345452  0.01627272
average_montly_hours  -0.52278654 -0.1795027  0.11612682
time_spend_company    -0.33143549  0.0645894 -0.43833959
Work_accident          0.06767655 -0.2622693 -0.09761638
left                  -0.22815071  0.6079660 -0.04655295
promotion_last_5years  0.01532004 -0.1605496 -0.63070885
salary                 0.04640023 -0.2692992 -0.53831188
```

PC1 21% **Work-Load/Ratings**	**PC2 16%** **Satisfaction**	**PC3 13%** **Compensation**
last_evaluation -0.46 number_project -0.55 average_monthly_hours -0.52	satisfaction_level -0.56 left +0.60	salary -0.53 time_spend_company -0.43 promotion_last_5years -0.63
Ratings go along with Number of projects and monthly efforts.	Left is inversely proportional to satisfaction	Salary and promotion are related to time spent in the company
Good rating employees usually overloaded and vice versa.	Attrition is mainly because of low satisfaction.	Salary is more related to time spent in the company rather than rating.

Conclusions

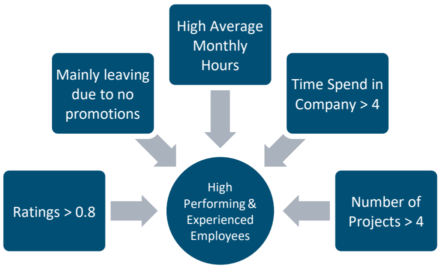
Why Are Employees Leaving?

From this report, it can be inferred that satisfaction is a significant factor for attrition. It is also observed that No. of Projects, Average Monthly Hours and Last Evaluation are the predictors of satisfaction level.

We recommendation below practices:

- Evaluation should not be purely focusing on the no. of projects/no. of average monthly hours of an employee.
- The workload of employees needs to be balanced.

Identify High Performing and Experienced Employees


High Average Monthly Hours
Mainly leaving due to no promotions
Time Spend in Company > 4
Ratings > 0.8
High Performing & Experienced Employees
Number of Projects > 4

Recommendation:

- Promotion/Performance Recognition plan needs to be devised for High Performing and Experienced Employees
- Average Monthly Hours of High Performing and Experienced Employees needs to be suitably reduced.

The scope of Further work

Other Factors to be considered for calculating *the Satisfaction* of an Employee

Work-Life Balance **|** Performance Recognition **|** Quality of Work **|** Improve Hiring Process **|**Capture Behavioural Metrics **|** Re-skilling **|** Training Need Analysis **|** Salary Benchmarking

These parameters will help to build a satisfaction model as mentioned in Create a Model for satisfaction Level***.***

Other Factors to be considered for calculating the *Real* Cost of losing an employee:

Hiring Cost**|** Onboarding Cost **|**Training Cost **|** Productivity/Operational Cost.

Risk Mitigation Steps

- **Control for Employee Burn-out**: Given that work hours seem to be one of the key factors influencing attrition, and steps need to be taken to reduce the no. of hours for overworked employees
- **Jobs with Opportunity to Grow**: Employees who are over-worked or under-work are could be dissatisfied due to no variation in the project they are working on. They should be provided with a variety of work.
- **Voluntary Attrition**-Attrition resulting from low ratings/evaluation may be good for the company. It would be good for a company to help them in upgrading their competencies with some training program.
- **Knowledge Management**: convert employees’ Tacit knowledge to Explicit knowledge.

Recommendations

- **Employee Development**: 'Empower and Grow' Employees. Install a proper training and competency development program.
- **Talent Acquisition**: 'Enhance' hiring process. Hire the right employees with market standard salaries.
- **Talent Movement Planning**: 'Groom' high potential employees. Give more responsibilities and better-quality work.
- **Best Practices**: Our initial study confirmed that Management department has the least attrition rate. It is advisable to get the best Practices from Management department.

Acknowledgments

We are pleased to acknowledge Prof. Ankur Sinha for his invaluable guidance, lectures during this project work. We extend our sincere thanks to our course faculty chair Prof. Arnab K. Laha who continuously motivated us throughout the course and project.

We also would like to thank all our other instructors (Prof. Karthik Sriram, Prof. Apratim Guha, Prof. Vineet Virmani, Prof. Kavitha Ranganathan, Prof. Goutam Dutta, and Prof. Sanjay Verma) whose lectures, notes and guidance have immensely helped us to complete our project.

Last but not the least, great thanks to our Programme Coordinator Ms. Jaya Nair, two research associates Mr. Rohit Bharadwaj and Ms. Pratishtha Batra and our fellow batchmates for their constant support throughout the course.

February 2019

Vikas Goel

Nilima Ghosh

Srikanth Shetty

References

1. Kaggle (https://www.kaggle.com) for the dataset
2. IIM Ahmedabad study material and codes provided during sessions.
3. Blogs on Data science algorithms – https://medium.com

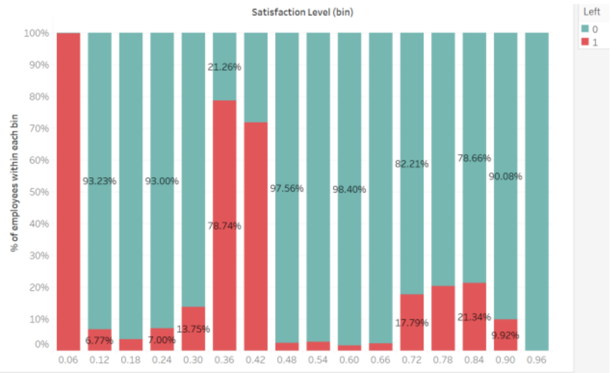
Appendix

EDA Plots

Observations through Data Visualization (using tableau)

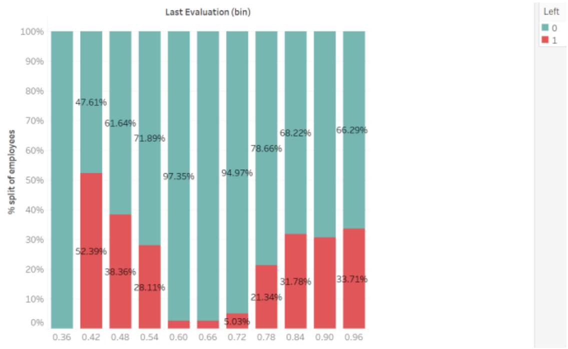
Note: - We have created bins in tableau for better analysis.

1. Satisfaction level



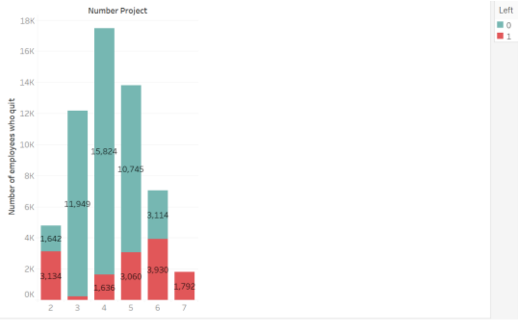
- Employees who had low satisfaction levels (0.2 or less) left the company
- employees who had low satisfaction levels (0.3~0.5) left the company more and
- employees who had high satisfaction levels (0.7 or more) left the company.

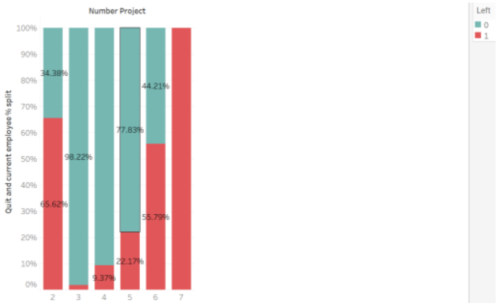
2. Last Evaluation



- Employees with Low evaluation (< 0.55) and with High Evaluation (>0.75) tend to leave the organization

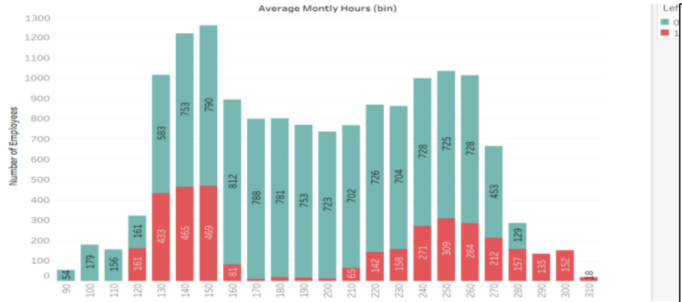
3. Number of projects versus the count of employees who leave



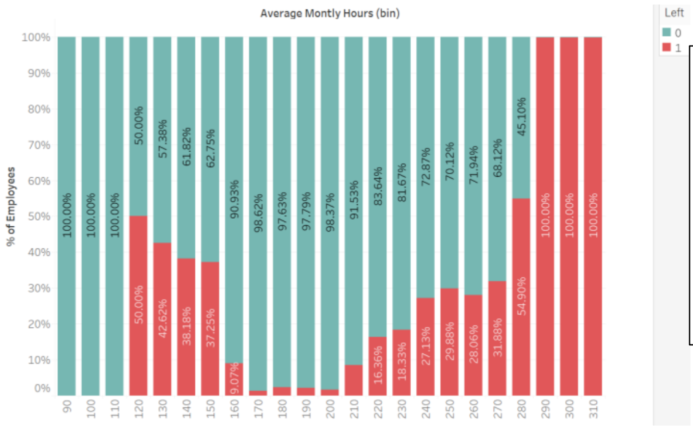


More than half of the employees with 2, 6, or 7 projects left the company. Majority of the employees who did not leave the company had 3,4, or 5 projects. All the employees with 7 projects left the company. There is an increase in employee turnover rate as project count increases.

4. Average monthly hours

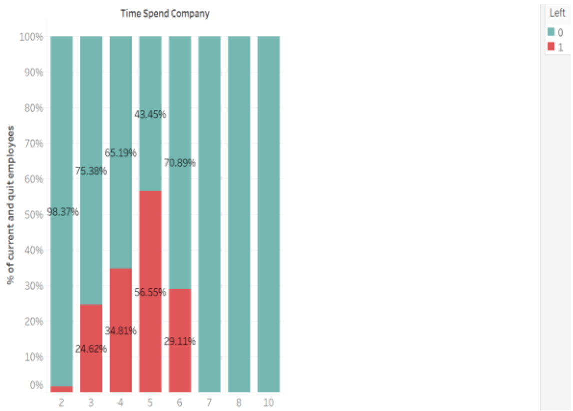


- The two extremes of average monthly hours might be area of interest. Employees who leave either work very high or low number of hours.
- Employee working between 160 – 220hour seems to be ideal to retain employees.



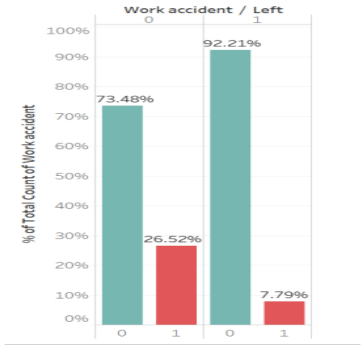
- This graph is inclusive of both active and left employees)
- Employees who had less hours of work (~150hours or less) left the company more. Employees who had too many hours of work (~250 or more) left the company

5. Time spent in the company



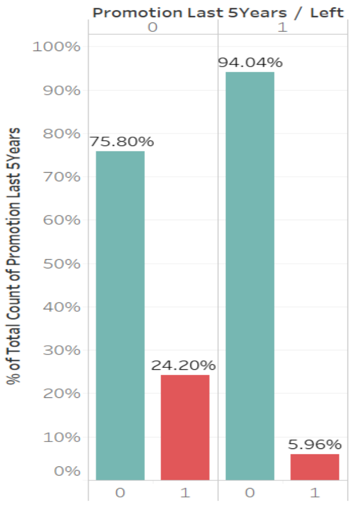
- The probability of employee leaving the company increases from 3 to 5.
- The attrition rate is highest for those employees who have spent 5 years at the company.
- People who have spent 2 years are not leaving the company.
- Once employees cross the golden years '7', they are not leaving.

6. Work Accident



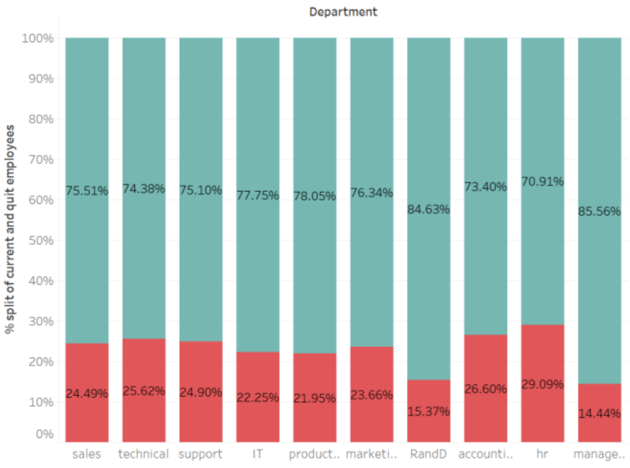
- The employees who had workplace accidents are less likely to leave than that of the employee who did not have workplace accidents

7. promotion_last_5years



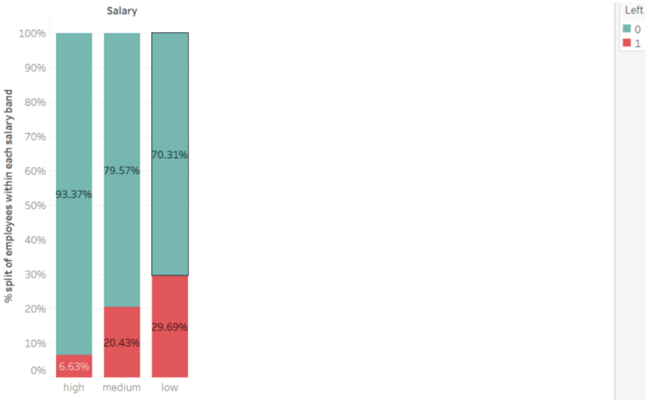
- The employees who were promoted in the last five years are less likely to leave than those who did not get a promotion in the last five years.

8. Department



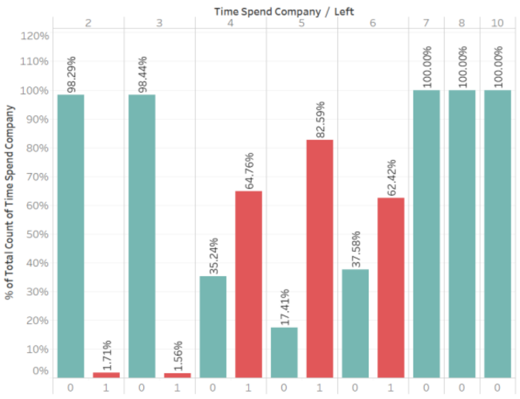
- HR department has highest attrition rate. Management department has least attrition rate did not get a promotion in the last five years.

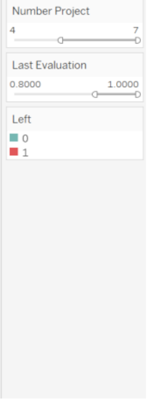
9. Salary



- The attrition within the group is highest in the salary group - low.
- The highest % of employees are in low salary bracket.

Why the “Best and Most Experienced Employees are Leaving”

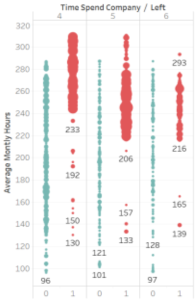


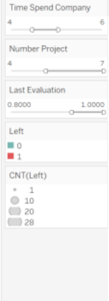


Defining Who Are the "Best “:- The median score is 0.72 -Anyone with 0.8 or higher scores are high performers.

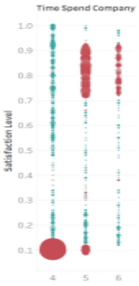
Defining Who are the "Most Experienced" -Number of projects>4

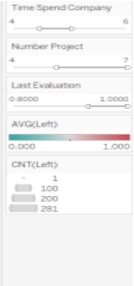
Reason 1 : Average Working Hours



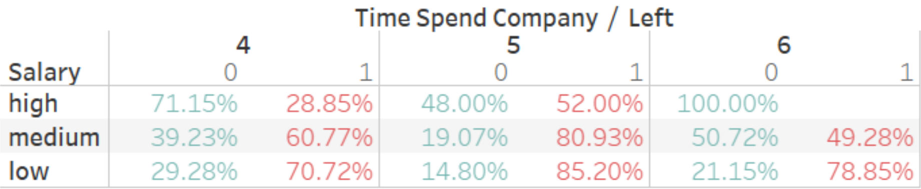


Reason 2 : Satisfaction Level



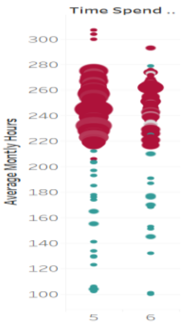


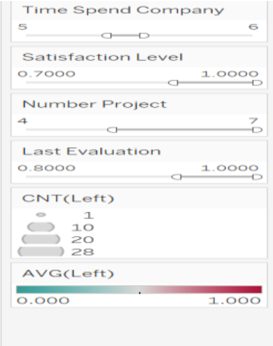
Reason 3 : Salary

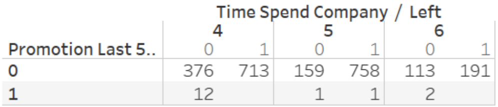


	Time Spend Company / Left					
	4		5		6	
Salary	0	1	0	1	0	1
high	71.15%	28.85%	48.00%	52.00%	100.00%	
medium	39.23%	60.77%	19.07%	80.93%	50.72%	49.28%
low	29.28%	70.72%	14.80%	85.20%	21.15%	78.85%

Why are ‘Best and Most Experienced Employees even with high Satisfaction levels Leaving?





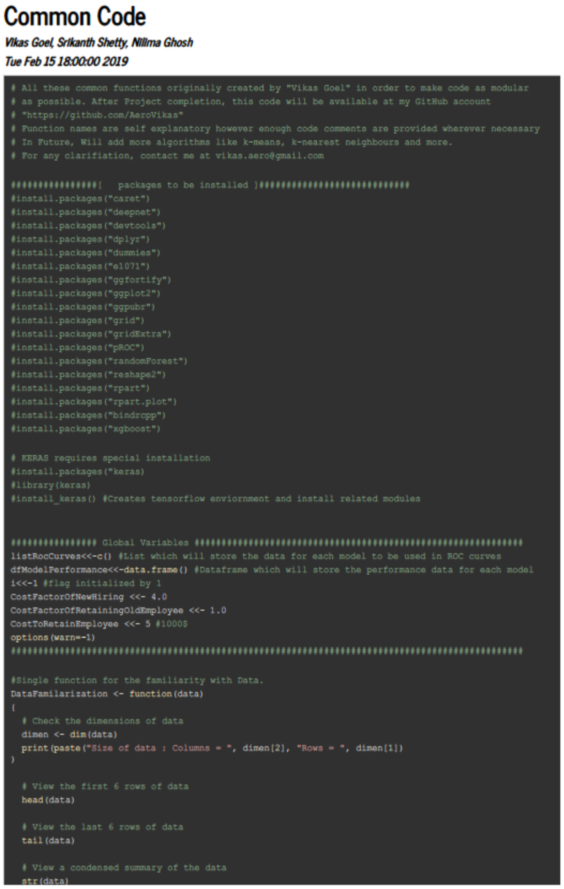
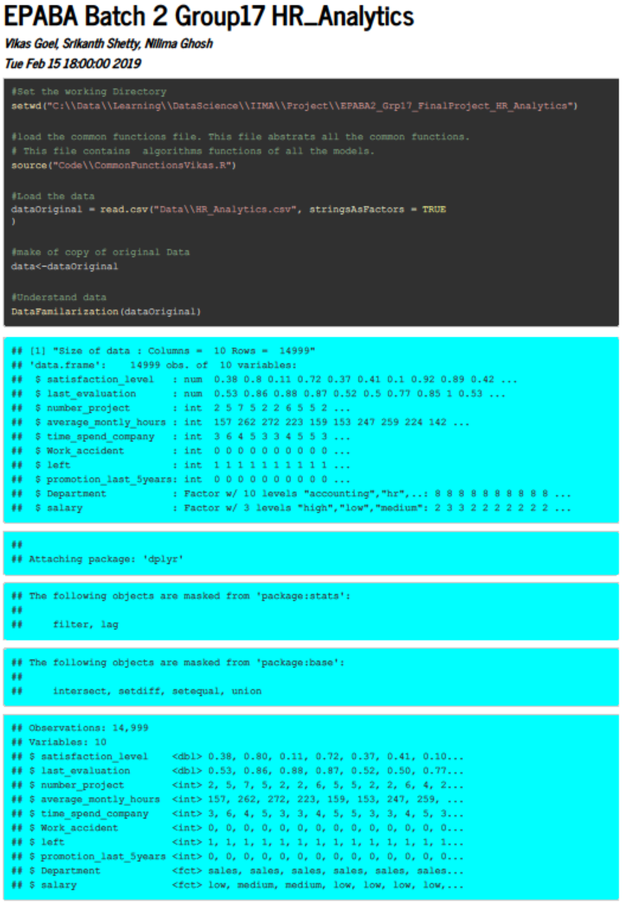


	Time Spend Company / Left					
	4		5		6	
Promotion Last 5..	0	1	0	1	0	1
0	376	713	159	758	113	191
1	12		1	1	2	

Very few of the ‘best and most experienced employees’ have received promotion in the last 5 years. Though there is no concrete evidence and striking difference to establish if not getting promoted in the last 5 years is the cause of employee leaving.

R Code and Results

Since we have a big piece of code, will provide R code and Result in pdf files as sperate zip as well as in this document.

	
CommonFunctionsVikas.pdf	EPABA2_Grp17_HR_Analytics Code and Results.pdf

Common Code

Vikas Goel, Srikanth Shetty, Nilima Ghosh

Tue Feb 15 18:00:00 2019

```
# All these common functions originally created by "Vikas Goel" in order to make code as modular
# as possible. After Project completion, this code will be available at my GitHub account
# "https://github.com/AeroVikas"
# Function names are self explanatory however enough code comments are provided wherever necessary
# In Future, Will add more algorithms like k-means, k-nearest neighbours and more.
# For any clarifiation, contact me at vikas.aero@gmail.com

################[   packages to be installed ]############################
#install.packages("caret")
#install.packages("deepnet")
#install.packages("devtools")
#install.packages("dplyr")
#install.packages("dummies")
#install.packages("e1071")
#install.packages("ggfortify")
#install.packages("ggplot2")
#install.packages("ggpubr")
#install.packages("grid")
#install.packages("gridExtra")
#install.packages("pROC")
#install.packages("randomForest")
#install.packages("reshape2")
#install.packages("rpart")
#install.packages("rpart.plot")
#install.packages("bindrcpp")
#install.packages("xgboost")

# KERAS requires special installation
#install.packages("keras)
#library(keras)
#install_keras() #Creates tensorflow enviornment and install related modules


################ Global Variables ############################################################
listRocCurves<<-c() #List which will store the data for each model to be used in ROC curves
dfModelPerformance<<-data.frame() #Dataframe which will store the performance data for each model
i<<-1 #flag initialized by 1
CostFactorOfNewHiring <<- 4.0
CostFactorOfRetainingOldEmployee <<- 1.0
CostToRetainEmployee <<- 5 #1000$
options(warn=-1)
##########################################################################################

#Single function for the familiarity with Data.
DataFamilarization <- function(data)
{
  # Check the dimensions of data
  dimen <- dim(data)
  print(paste("Size of data : Columns = ", dimen[2], "Rows = ", dimen[1]))
)

  # View the first 6 rows of data
  head(data)

  # View the last 6 rows of data
  tail(data)

  # View a condensed summary of the data
  str(data)
```

EPABA Batch 2 Group17 HR_Analytics

Vikas Goel, Srikanth Shetty, Nilima Ghosh

Tue Feb 15 18:00:00 2019

```
#Set the working Directory
setwd("C:\\Data\\Learning\\DataScience\\IIMA\\Project\\EPABA2_Grp17_FinalProject_HR_Analytics")

#load the common functions file. This file abstrats all the common functions.
# This file contains  algorithms functions of all the models.
source("Code\\CommonFunctionsVikas.R")

#Load the data
dataOriginal = read.csv("Data\\HR_Analytics.csv", stringsAsFactors = TRUE
)

#make of copy of original Data
data<-dataOriginal

#Understand data
DataFamilarization(dataOriginal)
```

```
## [1] "Size of data : Columns =  10 Rows =  14999"
## 'data.frame':    14999 obs. of  10 variables:
##  $ satisfaction_level   : num  0.38 0.8 0.11 0.72 0.37 0.41 0.1 0.92 0.89 0.42 ...
##  $ last_evaluation      : num  0.53 0.86 0.88 0.87 0.52 0.5 0.77 0.85 1 0.53 ...
##  $ number_project       : int  2 5 7 5 2 2 6 5 5 2 ...
##  $ average_montly_hours : int  157 262 272 223 159 153 247 259 224 142 ...
##  $ time_spend_company   : int  3 6 4 5 3 3 4 5 5 3 ...
##  $ Work_accident        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ left                 : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ promotion_last_5years: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Department           : Factor w/ 10 levels "accounting","hr",..: 8 8 8 8 8 8 8 8 8 8 ...
##  $ salary               : Factor w/ 3 levels "high","low","medium": 2 3 3 2 2 2 2 2 2 2 ...
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## Observations: 14,999
## Variables: 10
## $ satisfaction_level    <dbl> 0.38, 0.80, 0.11, 0.72, 0.37, 0.41, 0.10...
## $ last_evaluation       <dbl> 0.53, 0.86, 0.88, 0.87, 0.52, 0.50, 0.77...
## $ number_project        <int> 2, 5, 7, 5, 2, 2, 6, 5, 5, 2, 2, 6, 4, 2...
## $ average_montly_hours  <int> 157, 262, 272, 223, 159, 153, 247, 259, ...
## $ time_spend_company    <int> 3, 6, 4, 5, 3, 3, 4, 5, 5, 3, 3, 4, 5, 3...
## $ Work_accident         <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ left                  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1...
## $ promotion_last_5years <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0...
## $ Department            <fct> sales, sales, sales, sales, sales, sales...
## $ salary                <fct> low, medium, medium, low, low, low, low,...
```