HR dataset

* Satisfaction Level
* Last evaluation
* Number of projects
* Average monthly hours
* Time spent at the company
* Whether they have had a work accident
* Whether they have had a promotion in the last 5 years
* Departments (column sales)
* Salary
* Whether the employee has left

1. Develop a model which can predict whether a given employee will leave or not.

**Ans.**

## Model 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model  Number | Specificity Score | Sensitivity Score | Percentage of area under curve ROC | Accuracy | Column details which you considered to build a model | Any other columns that you wanted to add for comparison |
| 1 | 0.5256231 | 0.8866818 | 0.7062 | 80.072 | All columns except sales | sales |

## Summary:

Call:

glm(formula = left ~ salary + satisfaction\_level + number\_project +

time\_spend\_company + average\_montly\_hours + Work\_accident +

promotion\_last\_5years + last\_evaluation, family = "binomial",

data = dataset)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.2059 -0.6646 -0.4080 -0.1206 3.0918

Coefficients:

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

(Intercept) -1.5508132 0.1717124 -9.031 < 2e-16 \*\*\*

salarylow 2.0054295 0.1275116 15.727 < 2e-16 \*\*\*

salarymedium 1.4708288 0.1283351 11.461 < 2e-16 \*\*\*

satisfaction\_level -4.1307434 0.0977270 -42.268 < 2e-16 \*\*\*

number\_project -0.3149832 0.0212517 -14.822 < 2e-16 \*\*\*

time\_spend\_company 0.2612599 0.0153833 16.983 < 2e-16 \*\*\*

average\_montly\_hours 0.0044558 0.0005144 8.662 < 2e-16 \*\*\*

Work\_accident -1.5368679 0.0894349 -17.184 < 2e-16 \*\*\*

promotion\_last\_5years -1.4828491 0.2559823 -5.793 6.92e-09 \*\*\*

last\_evaluation 0.7264817 0.1486903 4.886 1.03e-06 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 16465 on 14998 degrees of freedom

Residual deviance: 12907 on 14989 degrees of freedom

AIC: 12927

Number of Fisher Scoring iterations: 5

## ROC Curve

## 

Inference:

From the above model, I selected all the features excluding the sales. As we can see that the area under the curve is 70% which means when we are going to predict whether a person will leave the company. The ROC curve does this by plotting sensitivity, the probability of predicting a real positive will be a positive, against 1-specificity, the probability of predicting a real negative will be a positive.

So, our model is able to distinguish between true positive and true negative values. The model accuracy is 80.72% percentage.

## Model 2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model  Number | Specificity Score | Sensitivity Score | Percentage of area under curve ROC | Accuracy | Column details which you considered to build a model | Any other columns that you wanted to add for comparison |
| 2 | 0.5329039 | 0.8858943 | 0.7094 | 80.18535 | All columns | - |

## Summary:

Call:

glm(formula = left ~ ., family = "binomial", data = dataset)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.2248 -0.6645 -0.4026 -0.1177 3.0688

Coefficients:

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

(Intercept) -1.4762862 0.1938373 -7.616 2.61e-14 \*\*\*

satisfaction\_level -4.1356889 0.0980538 -42.178 < 2e-16 \*\*\*

last\_evaluation 0.7309032 0.1491787 4.900 9.61e-07 \*\*\*

number\_project -0.3150787 0.0213248 -14.775 < 2e-16 \*\*\*

average\_montly\_hours 0.0044603 0.0005161 8.643 < 2e-16 \*\*\*

time\_spend\_company 0.2677537 0.0155736 17.193 < 2e-16 \*\*\*

Work\_accident -1.5298283 0.0895473 -17.084 < 2e-16 \*\*\*

promotion\_last\_5years -1.4301364 0.2574958 -5.554 2.79e-08 \*\*\*

saleshr 0.2323779 0.1313084 1.770 0.07678 .

salesIT -0.1807179 0.1221276 -1.480 0.13894

salesmanagement -0.4484236 0.1598254 -2.806 0.00502 \*\*

salesmarketing -0.0120882 0.1319304 -0.092 0.92700

salesproduct\_mng -0.1532529 0.1301538 -1.177 0.23901

salesRandD -0.5823659 0.1448848 -4.020 5.83e-05 \*\*\*

salessales -0.0387859 0.1024006 -0.379 0.70486

salessupport 0.0500251 0.1092834 0.458 0.64713

salestechnical 0.0701464 0.1065379 0.658 0.51027

salarylow 1.9440627 0.1286272 15.114 < 2e-16 \*\*\*

salarymedium 1.4132244 0.1293534 10.925 < 2e-16 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

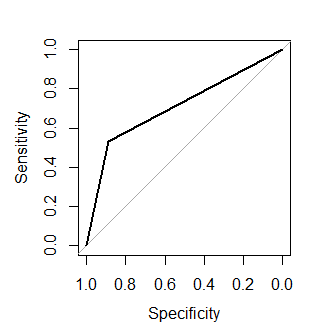
Null deviance: 16465 on 14998 degrees of freedom

Residual deviance: 12850 on 14980 degrees of freedom

AIC: 12888

Number of Fisher Scoring iterations: 5

## ROC Curve



Inference:

From the above model, I selected all the features. As we can see that the area under the curve is 70.9% which increased.

So, our model is able to distinguish between true positive and true negative values. But, our model accuracy decreased which is 80.18% percentage.

## Model 3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model  Number | Specificity Score | Sensitivity Score | Percentage of area under curve ROC | Accuracy | Column details which you considered to build a model | Any other columns that you wanted to add for comparison |
| 3 | 0.5035004 | 0.9048827 | 0.7042 | 80.93206 | Salary  +  Satisfaction\_level  +  Work\_accident  +  Promotion\_last\_5years  +  Last\_evaluation | Number\_project  Average\_montly\_hours  Time\_spend\_company  sales |

## Summary:

Call:

glm(formula = left ~ satisfaction\_level + last\_evaluation + Work\_accident +

promotion\_last\_5years + salary, family = "binomial", data = dataset)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.6665 -0.7007 -0.4520 -0.1595 3.1428

Coefficients:

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

(Intercept) -0.73791 0.15149 -4.871 1.11e-06 \*\*\*

satisfaction\_level -3.86846 0.09057 -42.712 < 2e-16 \*\*\*

last\_evaluation 0.54608 0.12357 4.419 9.91e-06 \*\*\*

Work\_accident -1.47626 0.08757 -16.858 < 2e-16 \*\*\*

promotion\_last\_5years -1.21188 0.24982 -4.851 1.23e-06 \*\*\*

salarylow 1.81241 0.12283 14.755 < 2e-16 \*\*\*

salarymedium 1.31424 0.12406 10.593 < 2e-16 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

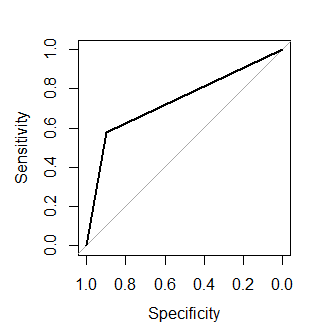
Null deviance: 16465 on 14998 degrees of freedom

Residual deviance: 13378 on 14992 degrees of freedom

AIC: 13392

Number of Fisher Scoring iterations: 5

## ROC Curve



Inference:

From the above model, I selected the following features Salary, Satisfaction\_level, Work\_accident, Promotion\_last\_5years, Last\_evaluation . As we can see that the area under the curve is 70.4% which decreased.

So, our model is able to distinguish between true positive and true negative values. The model accuracy is 80.9% percentage, as we can see that the accuracy percentage increased. Thus,we can observe that AUC value is decreasing while the accuracy value is increasing. It may mean that by selecting a particular feature we can increase the accuracy of our predicting model while this doesn’t means that the accuracy and AUC value will increase if we keep on adding the features in our model. If AUC value keep on increasing, it means our model is better at identifying the difference between classes.

## Model 4

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model  Number | Specificity Score | Sensitivity Score | Percentage of area under curve ROC | Accuracy | Column details which you considered to build a model | Any other columns that you wanted to add for comparison |
| 4 | 0.5838701 | 0.8983199 | 0.7411 | 82.34549 | Salary  +  Satisfaction\_level+  Number\_project  +  Work\_accident  +  Promotion\_last\_5years  +  sales | Average\_montly\_hours  Time\_spend\_company  Last\_evaluation |

Summary

Call:

glm(formula = left ~ salary + satisfaction\_level + number\_project +

promotion\_last\_5years + Work\_accident + sales, family = "binomial",

data = dataset)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.7913 -0.6932 -0.4414 -0.1506 3.1927

Coefficients:

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

(Intercept) 0.19408 0.16957 1.144 0.2524

salarylow 1.77119 0.12340 14.353 < 2e-16 \*\*\*

salarymedium 1.28697 0.12451 10.336 < 2e-16 \*\*\*

satisfaction\_level -3.97548 0.09293 -42.778 < 2e-16 \*\*\*

number\_project -0.10916 0.01614 -6.765 1.33e-11 \*\*\*

promotion\_last\_5years -1.18841 0.25052 -4.744 2.10e-06 \*\*\*

Work\_accident -1.45391 0.08727 -16.660 < 2e-16 \*\*\*

saleshr 0.18955 0.12910 1.468 0.1420

salesIT -0.17073 0.11960 -1.427 0.1534

salesmanagement -0.32054 0.15790 -2.030 0.0423 \*

salesmarketing 0.02093 0.12915 0.162 0.8712

salesproduct\_mng -0.16130 0.12788 -1.261 0.2072

salesRandD -0.59145 0.14146 -4.181 2.90e-05 \*\*\*

salessales -0.03397 0.10037 -0.338 0.7351

salessupport 0.01545 0.10713 0.144 0.8853

salestechnical 0.03181 0.10434 0.305 0.7605

---

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

(Dispersion parameter for binomial family taken to be 1)

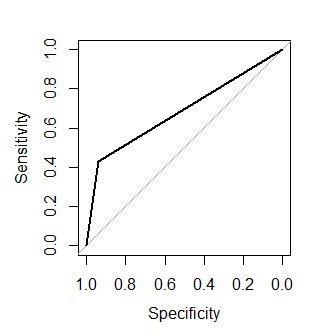
Null deviance: 16465 on 14998 degrees of freedom

Residual deviance: 13301 on 14983 degrees of freedom

AIC: 13333

Number of Fisher Scoring iterations: 5

ROC



Inference:

From the above model, I selected all the features excluding the sales. As we can see that the area under the curve is 74.11% which means that this model is better at distinguishing the classes from one another. In this model larger area lie under ROC in comparison to others.

The model accuracy is 82.34% percentage. In this model the accuracy of our model is increased along with that the AUC value is also increased which means that our model is better at predicting the response variable.

2. Prepare a table which compares with below values:

1. Model Number
2. Specificity Score
3. Sensitivity Score
4. Percentage of area under curve ROC
5. Accuracy
6. Column details which you considered to build a model
7. Any other columns that you wanted to add for comparison

**Ans.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model  Number | Specificity Score | Sensitivity Score | Percentage of area under curve ROC | Accuracy | Column details which you considered to build a model | Any other columns that you wanted to add for comparison |
| 1 | 0.5256231 | 0.8866818 | 0.7062 | 80.072 | Salary  +  Satisfaction\_level+  Number\_project+  Time\_spend\_company  +  Average\_montly\_hours  +  Work\_accident  +  Promotion\_last\_5years  +  last\_evaluation | sales |
| 2 | 0.5329039 | 0.8858943 | 0.7094 | 80.18535 | Salary  +  Satisfaction\_level+  Number\_project  +  Average\_montly\_hours  +  Work\_accident  +  Promotion\_last\_5years  +  Last\_evaluation  +  Time\_spend\_company  +  sales | - |
| 3 | 0.5035004 | 0.9048827 | 0.7042 | 80.93206 | Salary  +  Satisfaction\_level  +  Work\_accident  +  Promotion\_last\_5years  +  Last\_evaluation | Number\_project  Average\_montly\_hours  Time\_spend\_company  sales |
| 4 | 0.5838701 | 0.8983199 | 0.7411 | 82.34549 | Salary  +  Satisfaction\_level+  Number\_project  +  Work\_accident  +  Promotion\_last\_5years  +  sales | Average\_montly\_hours  Time\_spend\_company  Last\_evaluation |

3. Finally state the model of your choice and describe the reason for your model preference.

**Ans.**

In my opinion the model number 4 is giving more accuracy due to the features I selected. The model of my choice is this model because in this model, we found that dependant variable left is dependant on the features like salary, satisfaction level, number of projects, work accidents, promotion in the last five years and types of job in sales. In my opinion that the choosing of the features for this model is more accurately predict the response variable. A person who is more satisfied will also dependant on different variables such as time spent in the company, average monthly hours. So, I have not taken those features in my model building. And another feature which is last evaluation, which is also not the right feature to predict the person is going to leave the company or not.

Thus, after analyzing the features given in the dataset, I have taken only those features. As if a person is satisfied with his/her job he will not leave the job, or any exception. Similarly, if a person is not getting a promotion in the last five years, s/he will get demotivated. Other features like salary also play a major role in deciding our response variable. Job type also determines whether the person is going to leave the job or not.

Hence, in my opinion the model number 4 will give the best prediction.

# EDA

## Objective

The objective of the assignment is to use statistical techniques for the given dataset HR and find the reason behind the leaving of the employees. For that purpose, need to build a model which will help to predict either an employee will stay in the company or leave it. The main aim is to build model which will best predict that employees will leave the company or not.

## Introduction

In a company the predicting whether an employee will leave the company or not will depend on many factors. As the reason may be different for different for different person. The task is to build a model which will help to predict this based on several other reasons. To avoid the leaving of the employees, the model will help to predict.

## Data for analysis

There 10 variables and 14999 observations are present. Following are the variables present in the dataset:

* dddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddddd

10 Variables 14999 Observations

-----------------------------------

satisfaction\_level

n missing distinct

14999 0 92

Info Mean Gmd (Gini's mean difference)

1 0.6128 0.2823

.05 .10 .25

0.11 0.21 0.44

.50 .75 .90

0.64 0.82 0.92

.95

0.96

lowest : 0.09 0.10 0.11 0.12 0.13

highest: 0.96 0.97 0.98 0.99 1.00

-----------------------------------

last\_evaluation

n missing distinct

14999 0 65

Info Mean Gmd

1 0.7161 0.1973

.05 .10 .25

0.46 0.49 0.56

.50 .75 .90

0.72 0.87 0.95

.95

0.98

lowest : 0.36 0.37 0.38 0.39 0.40

highest: 0.96 0.97 0.98 0.99 1.00

-----------------------------------

number\_project

n missing distinct

14999 0 6

Info Mean Gmd

0.945 3.803 1.367

Value 2 3 4 5

Frequency 2388 4055 4365 2761

Proportion 0.159 0.270 0.291 0.184

Value 6 7

Frequency 1174 256

Proportion 0.078 0.017

-----------------------------------

average\_montly\_hours

n missing distinct

14999 0 215

Info Mean Gmd

1 201.1 57.48

.05 .10 .25

130 137 156

.50 .75 .90

200 245 267

.95

275

lowest : 96 97 98 99 100

highest: 306 307 308 309 310

-----------------------------------

time\_spend\_company

n missing distinct

14999 0 8

Info Mean Gmd

0.905 3.498 1.43

Value 2 3 4 5

Frequency 3244 6443 2557 1473

Proportion 0.216 0.430 0.170 0.098

Value 6 7 8 10

Frequency 718 188 162 214

Proportion 0.048 0.013 0.011 0.014

-----------------------------------

Work\_accident

n missing distinct

14999 0 2

Info Sum Mean

0.371 2169 0.1446

Gmd

0.2474

-----------------------------------

left

n missing distinct

14999 0 2

Info Sum Mean

0.544 3571 0.2381

Gmd

0.3628

-----------------------------------

promotion\_last\_5years

n missing distinct

14999 0 2

Info Sum Mean

0.062 319 0.02127

Gmd

0.04163

-----------------------------------

sales

n missing distinct

14999 0 10

accounting (767, 0.051), hr (739,

0.049), IT (1227, 0.082),

management (630, 0.042), marketing

(858, 0.057), product\_mng (902,

0.060), RandD (787, 0.052), sales

(4140, 0.276), support (2229,

0.149), technical (2720, 0.181)

-----------------------------------

salary

n missing distinct

14999 0 3

Value high low medium

Frequency 1237 7316 6446

Proportion 0.082 0.488 0.430

-----------------------------------

## Exploratory Data Analysis

str(dataset)

'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 ...

$ sales : 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 ...

## Univariate Analysis

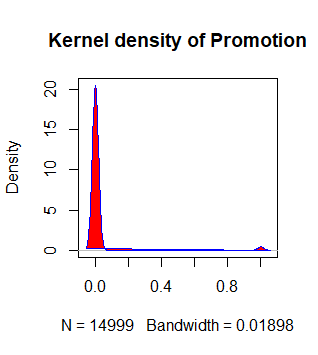


Fig: Kernel density plot of promotion

From the above plot, we can conclude that there are very less number of promotions in the company. As the plot is peaked near the region of 0.0 and there is very less density around 0.9.

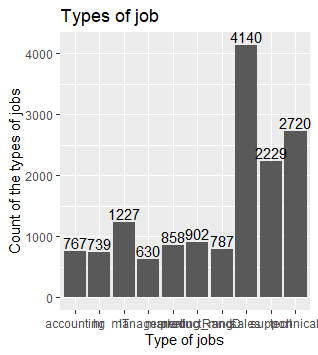


Fig: Types of jobs in the company and their count

From the above, I conclude that sales job is the job where many number of people work. So, top three types of job where larger number of people work are sales, support and technical.

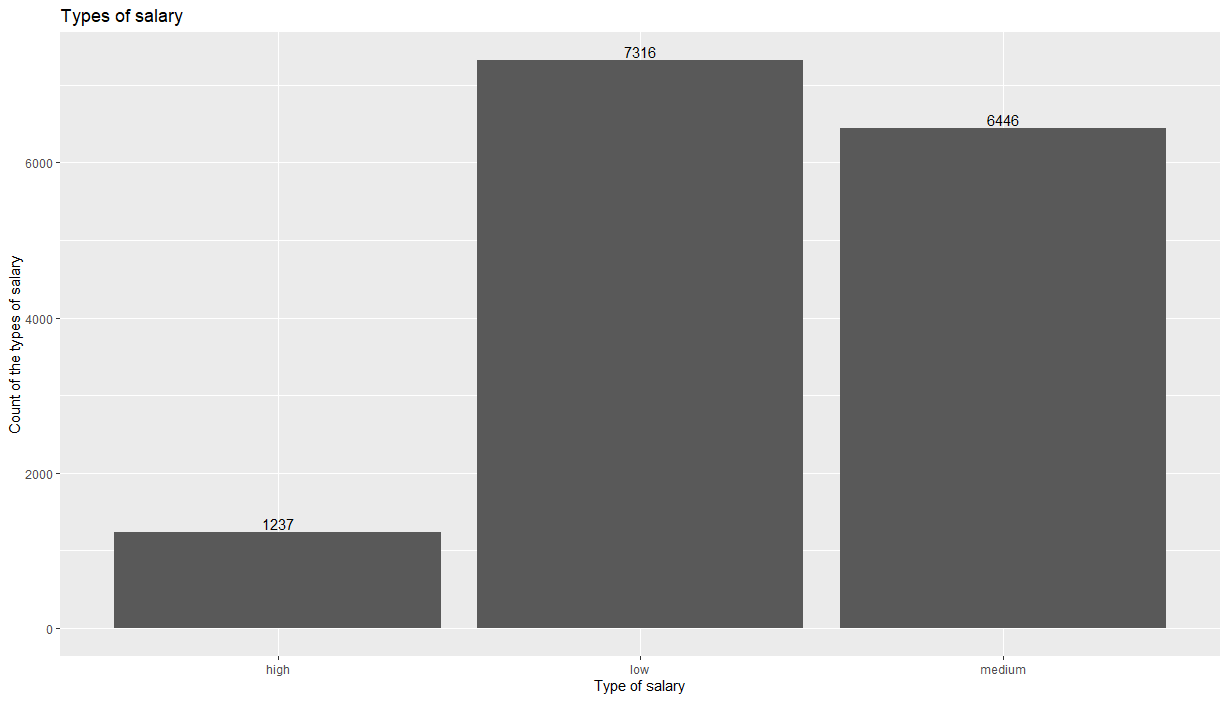


Fig: Count of the types of salary given to the employee

From the above plot, I can conclude that larger number of people are working under low salary and very less number are paid high by the company.

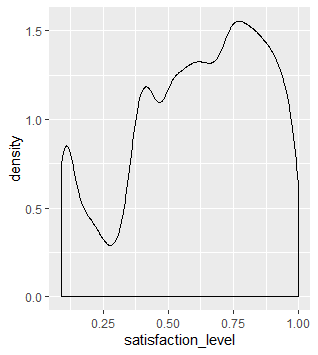


Fig: Density plot of satisfaction level

From the above, I can conclude that more number of people are satisfied and also there is fall near 0.25 to 0.75, which means there are a group of people who are not satisfied with their job.

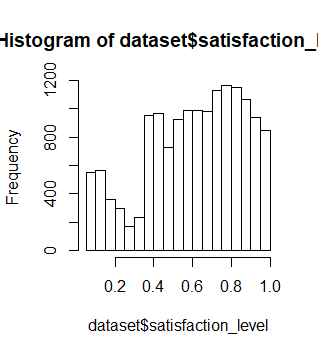


Fig: Histogram plot of Satisfaction level

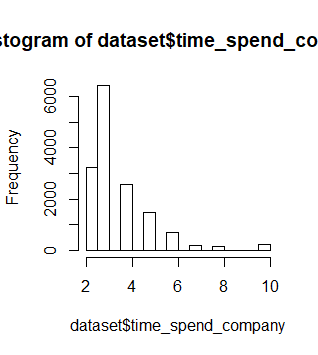


Fig: Histogram plot of time spend in the company

From the above plot, I can infer that a large number of people are spending less time in the company. Thus, the most common time spend in the company is in between 2 and 4.

## Bivariate/Multivariate Analysis

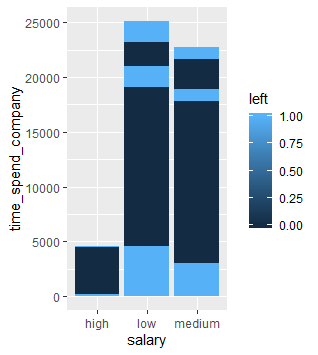


Fig: Relation between salary, time spent in the company and people left.

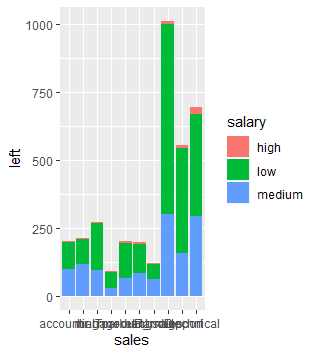


Fig: Relation between sales, left and salary

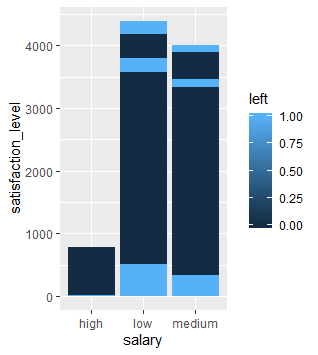


Fig: Relation between salary, satisfaction level and left

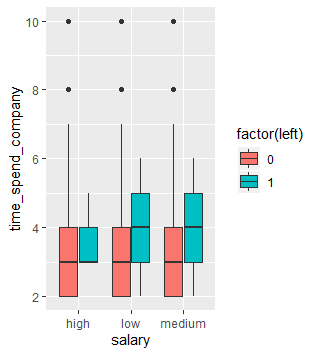


Fig: Relation between salary, time spend in company and left

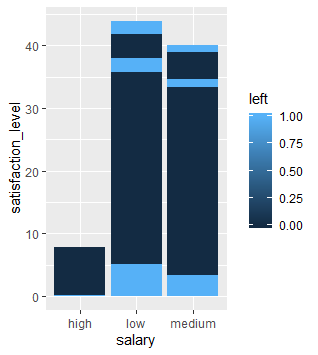


Fig: Relation between salary, satisfaction level and left

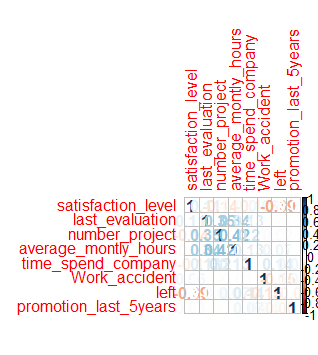


Fig: Correlation table

In this table we can see that there are many features which are not positively or negatively related.