**ARTIFICIAL INTELLIGENCE-ITE2010**

**REVIEW-2**

**FACULTY: PROF AJIT KUMAR SANTRA**

**SLOT-A2+TA2**

**TEAM**

**ANISH SINGH WALIA- REG-No:15BIT0116**

**VINAY-15BIT0165**

**KRISHNA YADAV-15BIT0088**

**PROJECT TITLE**-HUMAN RESOURCE ANALYTICS

**SUMMARY**- Review 2 will contain mostly about data preparation, data preprocessing, descriptive data analytics (what is happening?) using various statistical and data visualization techniques, analyzing the human resources data set, doing multi-variate, bi-variate and univariate analysis. Then using the knowledge extracted and inferred from descriptive analytics to do predictive modelling and model building and designing the architecture of the model.

**1)Dataset -**Human resource dataset. The dataset has 10 attributes(variables) and 14999 observations(instances).

**data.frame': 14999 obs. of 10 variables:**

**ATTRIBUTES** **TYPE** **INSTANCES**

$ 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 ...

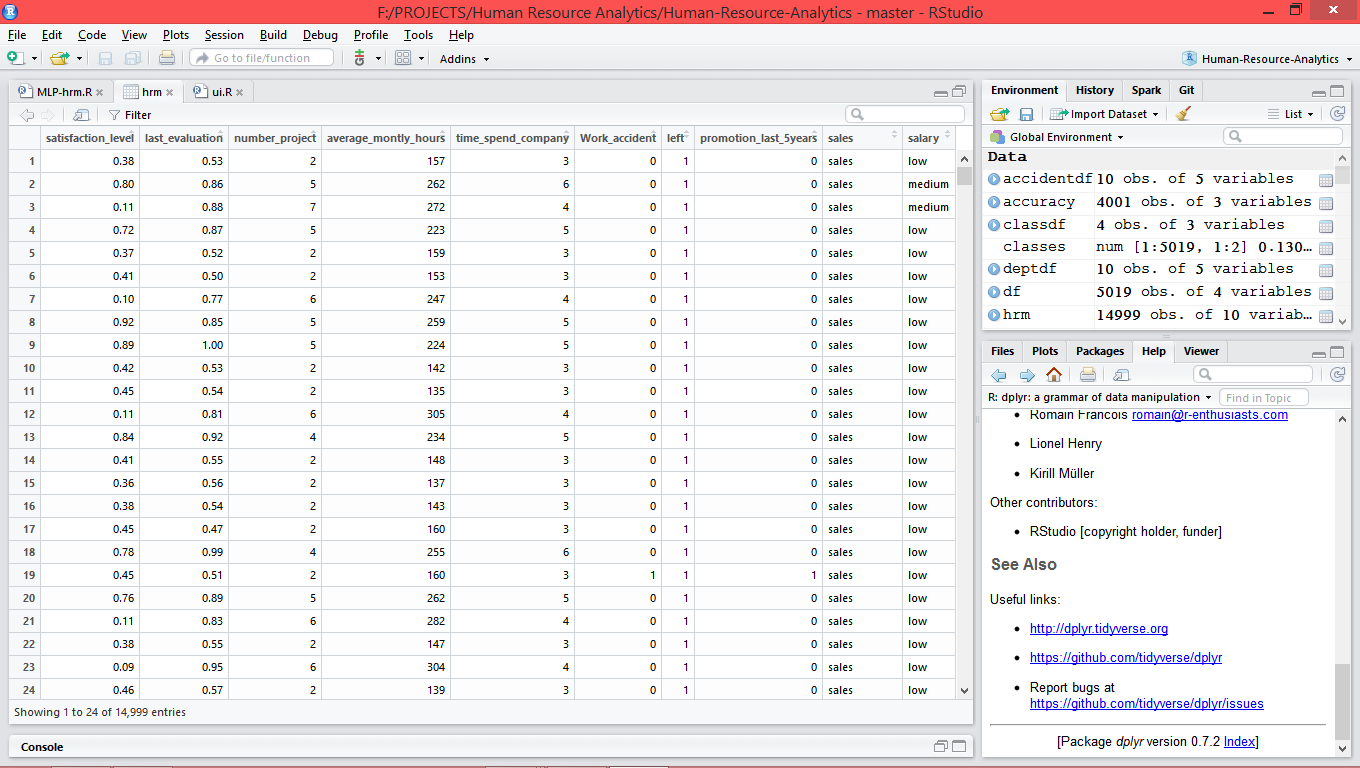
$ left : Factor w/ 2 levels 1 1 1 1 1 1 1 1 1 1 ...

$ 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 ...

**Sample Instances**:



*DATA PRE-PROCESSING*

1)**Data preprocessing** is done so as to make data suitable to feed it to a neural network.

2)Random sampling of data is done- i.e it is shuffled.

3)Normalization operations and scaling numeric variables to range [0,1] to reduce the computational complexity for the model.

#Min-max normalization function in R

normal<-function(x)

+ {

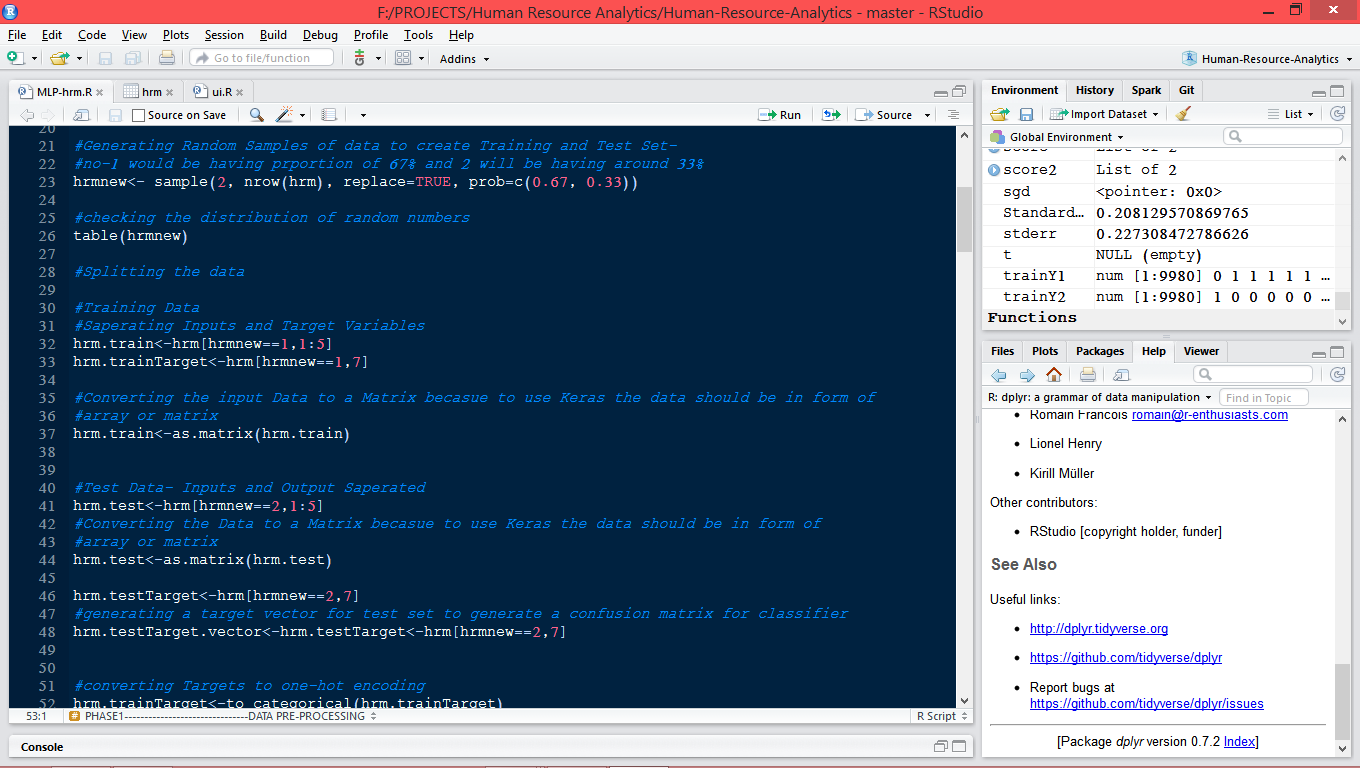
+ norm\_x<-(x-min(x))/(max(x)-min(x))

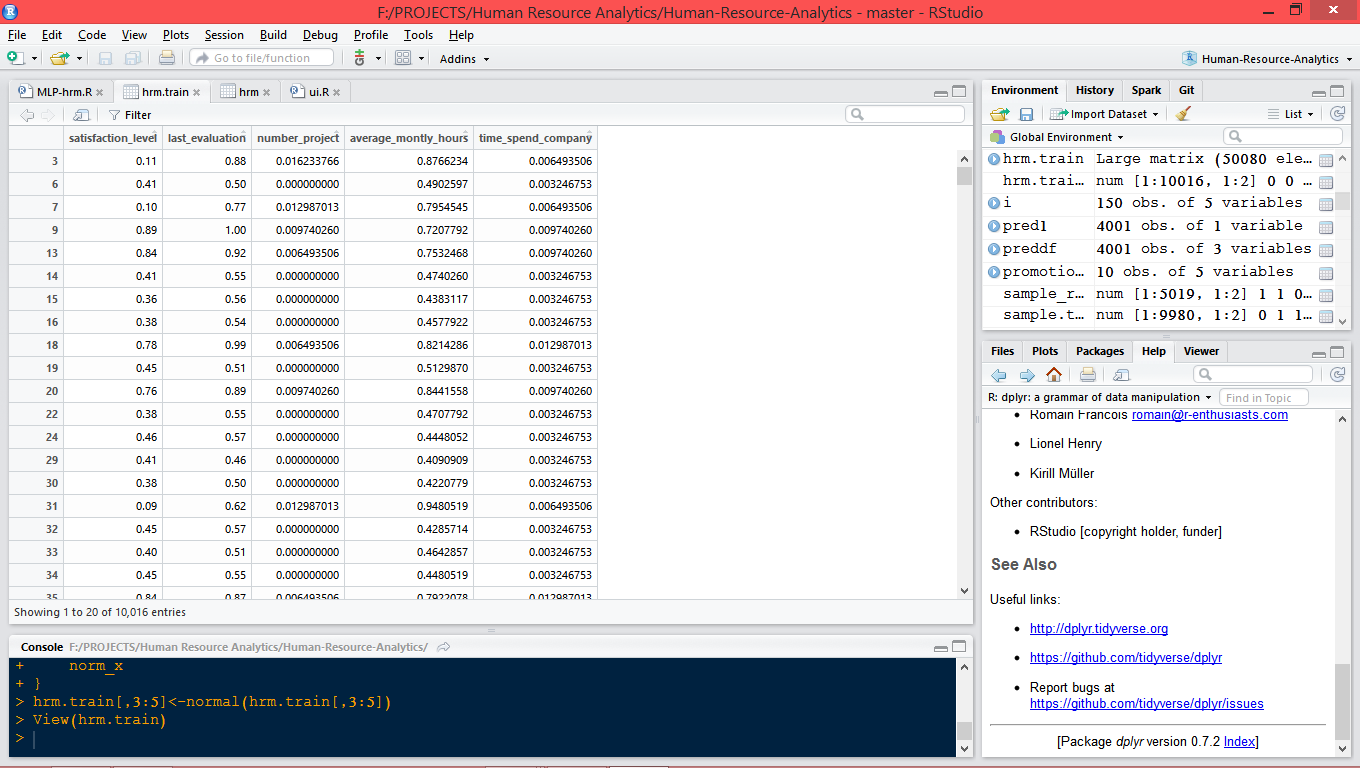
+ norm\_x

+ }

4)Separating Training and test data and selecting 5 best variables which explain and summarize most information in dataset.

5)Converting the data into **matrix** form so as to feed it to a neural network.





**Normalized training data-i.e between range [0,1]**

**3) Training a Multi-layer Deep Neural network**

**Problem Statement-** Predicting whether an employee will *leave the company* or not based on five inputs after basic data analytics and exploring the dataset.

MLP architecture-

Training inputs instances:9980 observations

Layers in Neural network-4 layers

**Input Units** – 5 input neurons as we have 5 input variables

**Hidden layer 1** – 32 hidden neurons

**Hidden layer 2** – 10 hidden neurons

**Output Units** — 2 output neurons (as we have 2 class labels)

The activation function used in order to learn non-linear interactions by the *hidden layers* are **ReLu**-rectified linear units and this function is of form- , i.e it converts negative values to 0.

Diagram of the architecture--

**Layer (type) Output Shape Param #**

=====================================================================================================

dense\_1 (Dense) (None, 32) 192

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_2 (Dense) (None, 10) 330

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_3 (Dense) (None, 2) 22

=====================================================================================================

Total params: 544

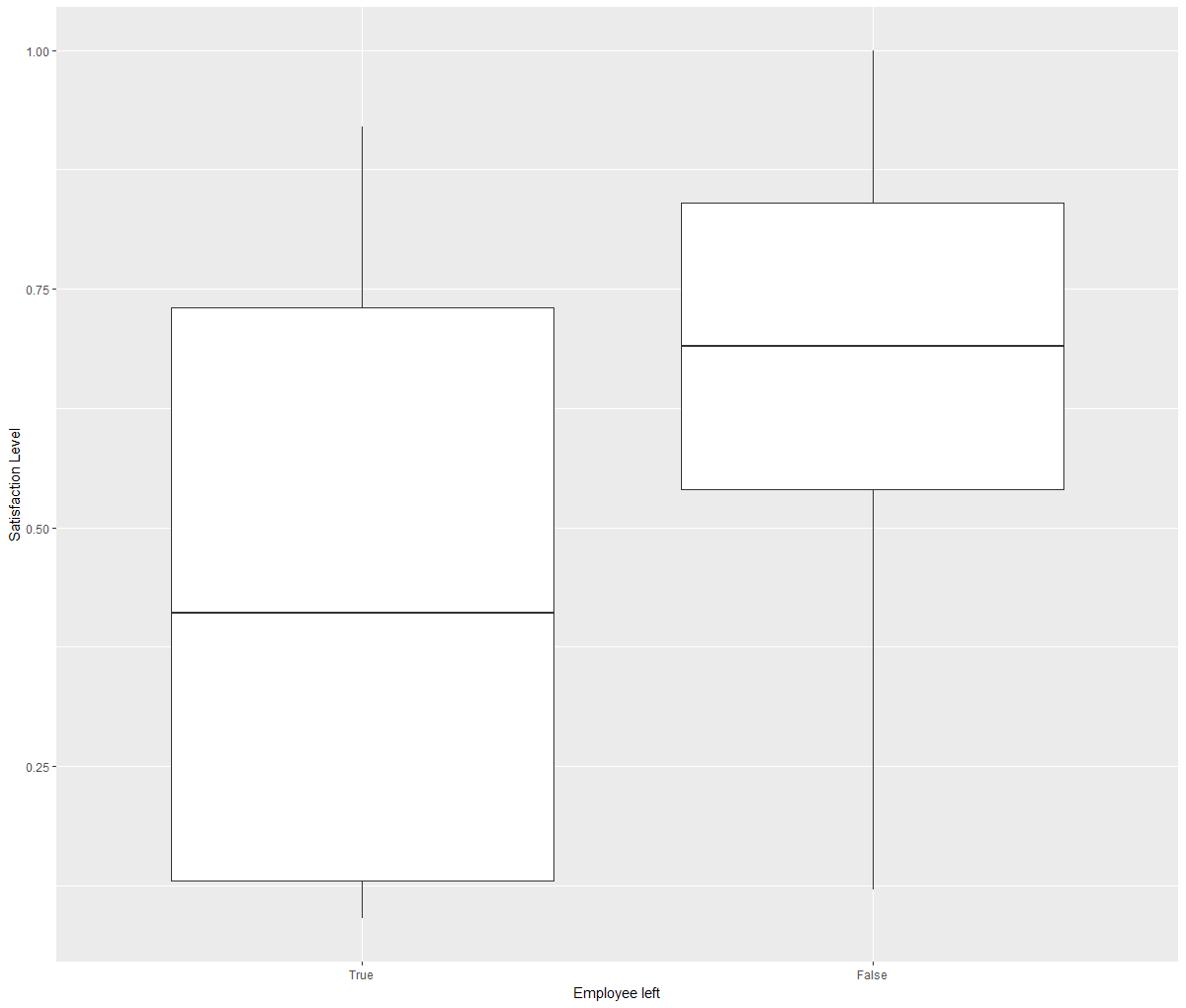
Trainable params: 544

Non-trainable params: 0

Hence the number of trainable parameters are **544 i.e weights and bias** as they are learned and tuned during the learning process of the model from training data.

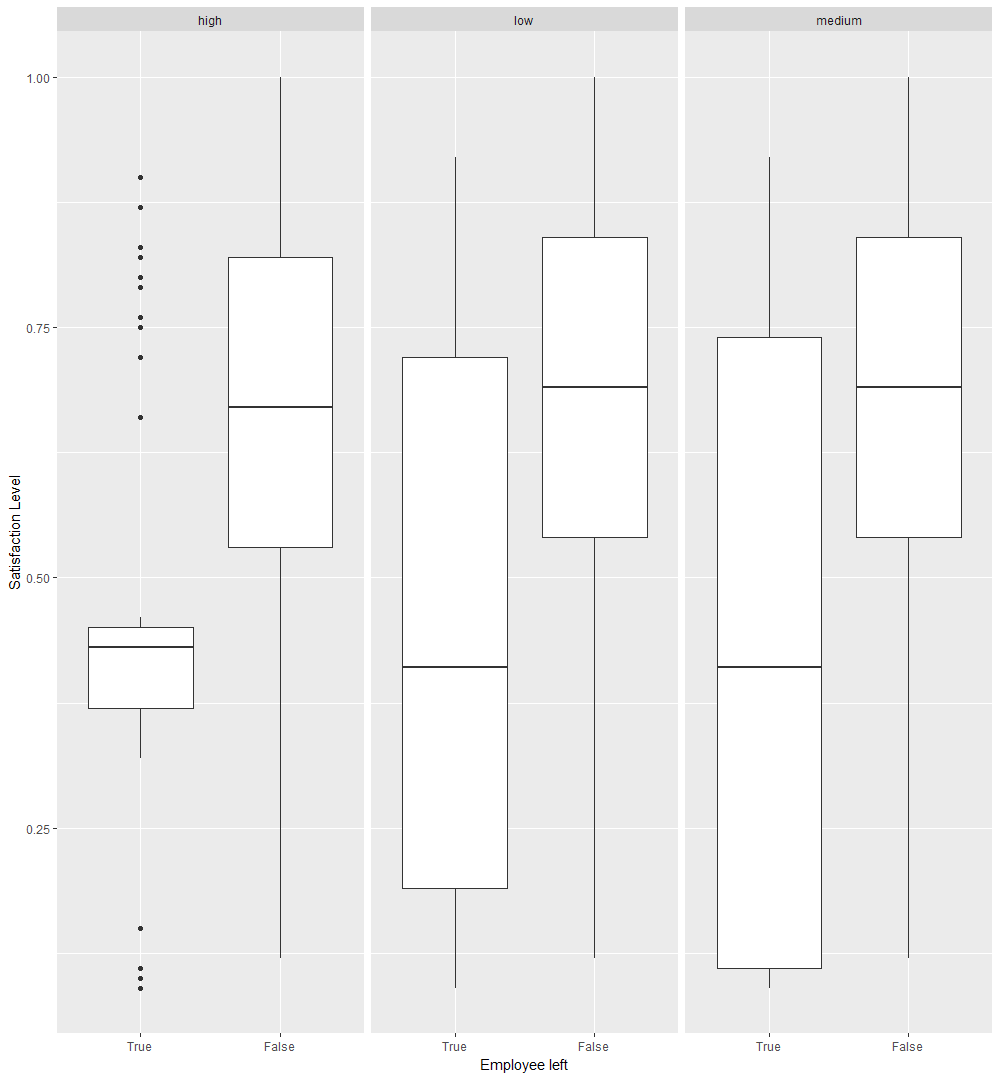
4)**DESCRIPTIVE ANALYSIS** of dataset

BOX PLOTS obtained while analyzing the dataset.

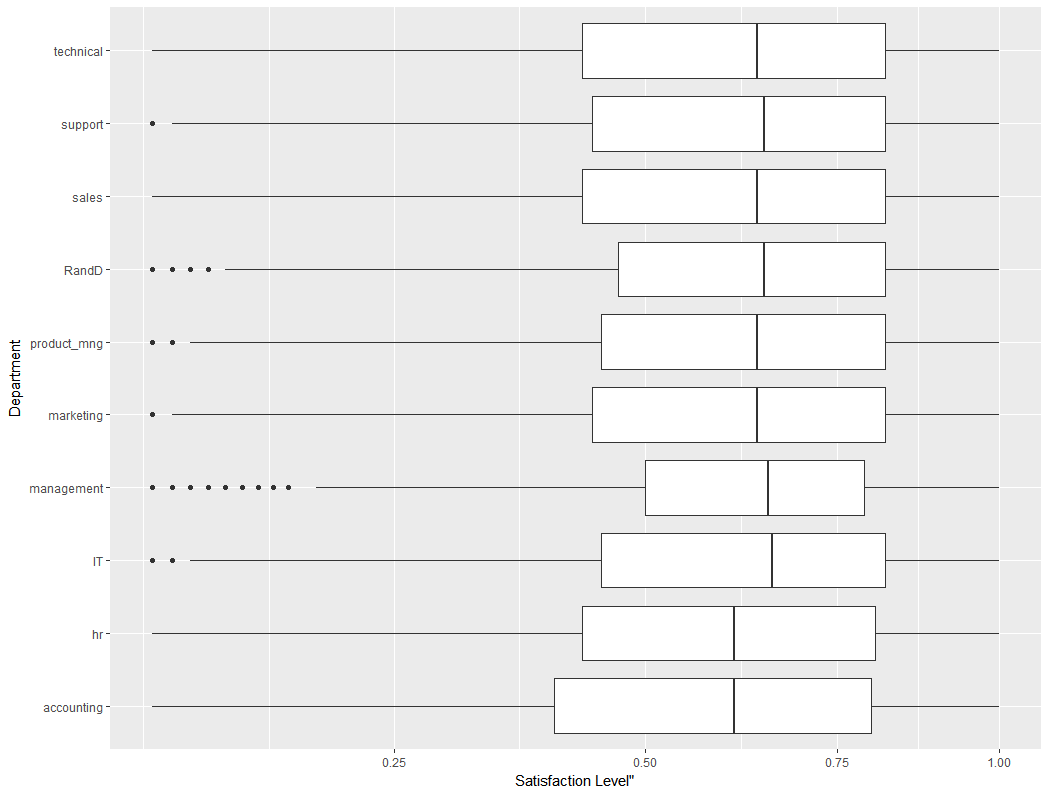


The plot above shows the boxplot which explains the statistical parameters such as median, quantiles and percentiles of a categorical variable and a numeric variable.

In the above plot we can see that employee who left have lower median satisfaction levels of around 0.35. Employee who didn’t leave the company have higher median satisfaction levels of around 0.60.



This boxplot shows the multi-variate analysis of three different variables- Employee\_left, Satisfaction level and salary.



Boxplot of satisfaction level of employees department wise

*One can notice that the highest satisfaction level is for employees in IT, management and R & D departments.*