**Predictive Modelling - OPIM 5604**

**HR Employee Attrition**



**INSTRUCTOR: Group 2**

Jose CRUZ Natasha Arora

Ishan sinha

JASLEEN KAUR

RAVI GURRAM

**Contents**

Objective 3

1.1 Introduction 4

[Data](#_Data_Overview:) Exploration 5

2.1 Exploration 5

2.2 Correlations 13

Data Modification 14

3.1 Exploring Missing Values14

3.2 Outlier Analysis 15

3.3 Data Reduction 16

[Data](#_Dimension/Variable_Reduction) Modeling 17

[4.1](#_Feature_Analysis) Model Building 17

Conclusions 24

[Recommendations](#_Data_sampling_(partition) 25

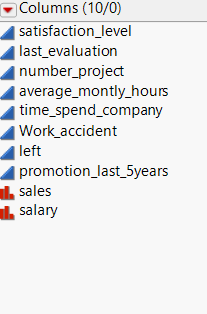
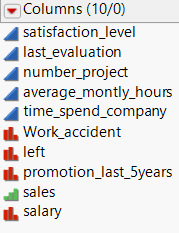
**Objective**

**Turnover Rate: 23.8%**

**Mean Satisfaction of employees: 0.61**

**Goal:** To predict which employees might leave based on patterns that might be presented in existing data.

**Variable Transformation:** Changed target variable to Nominal and variables Work\_accident, promotion\_last\_5\_years are also converted to Nominal with salary to Ordinal.

**1.1 Introduction**

Business Case:

A company never like to loose their loyal employees. In a way, if we look at it , it’s the employees that go a long way in making the company a huge success and intern they expect that the compant should reward them based on their work and behaviour. Not in every possible scenario , this could happen, because it’s not only the employees that make a firm, rather the cultural values,leadership team and several other factors go into picture which making a firm into a success.

But there are certain scenarios, where a employee is not satisfied with the firm and it results in frustration and eventually leaving the firm. We have pulled up a data set from Kaggle to better understand the underlying reasons so as to why employees leave the firm and have tried to give recommendations based on our analysis.

Employee leaving a firm leaves a great impact on the firm and it’s associated business value:

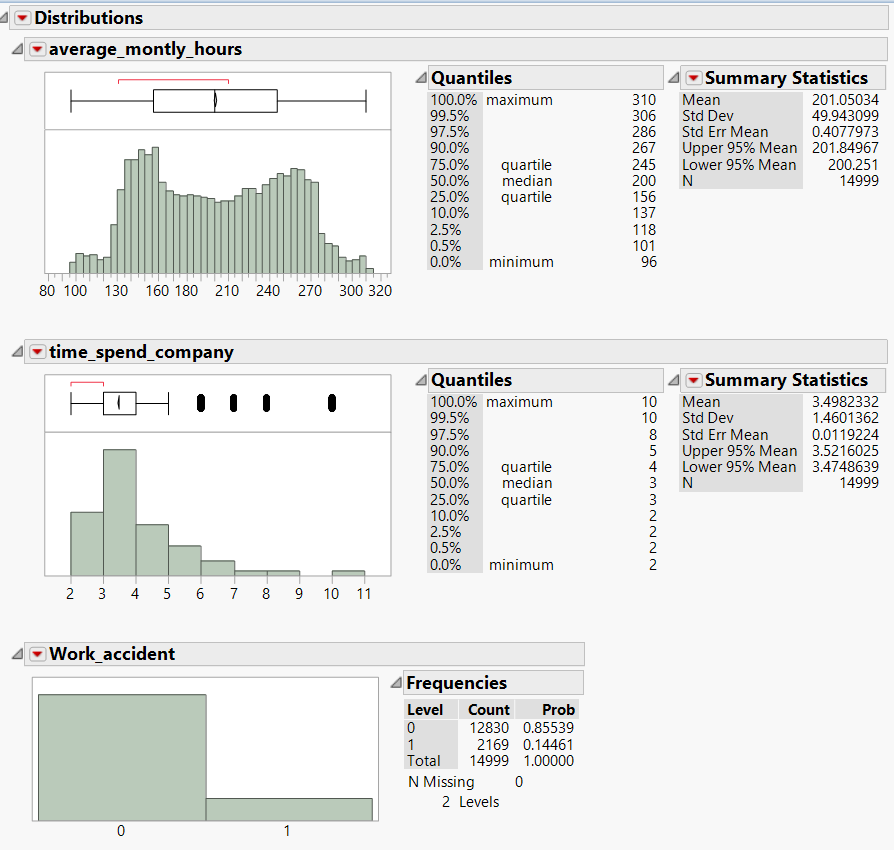
* Brand of the firm: When multiple employees leave the firm, it’s the brand value of the firm which gets degraded as in the market, the talks of employee leaving the firm creates abuzz and hence hampering the business value
* Reduction in Quality Manpower: When an employee leaves a firm, a good employee goes, because he or she is getting paid better in the outside market for the work that he/she is doing in your company, hence it impacts the firm/department’s deliverables as well.
* Impact on Revenue: A firm invests a lot of money in hiring and training the new talent and when that employee leaves the firm, the loss is beared by the company only. Consider this situation for many employees and imagine the amount lost in it.

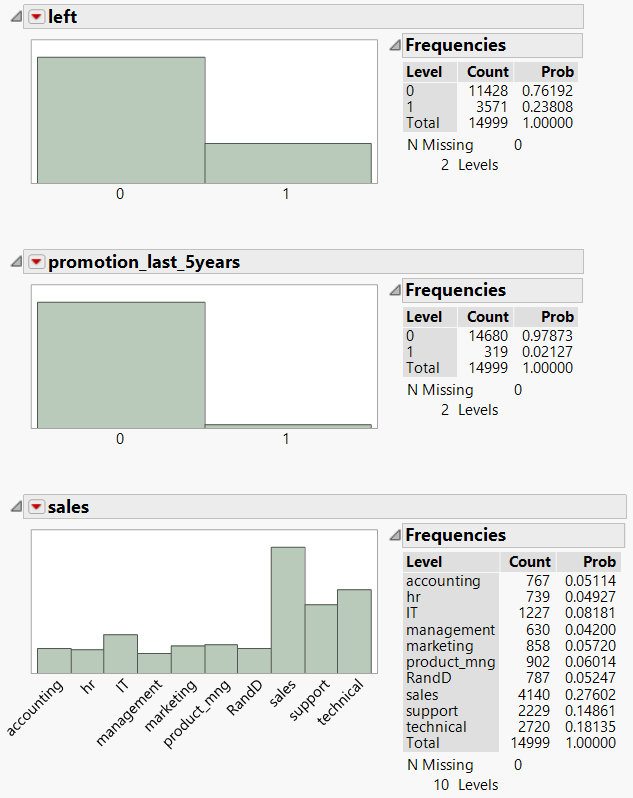
**Data Exploration**

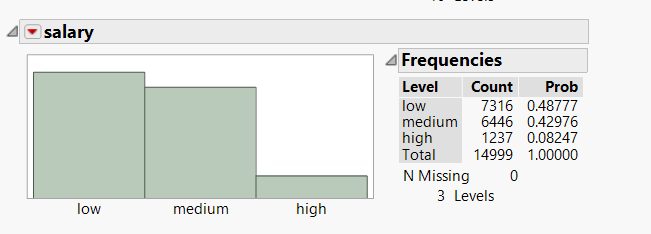
**2.2 Exploration:**

1. **Exploring distribution of each variable:**







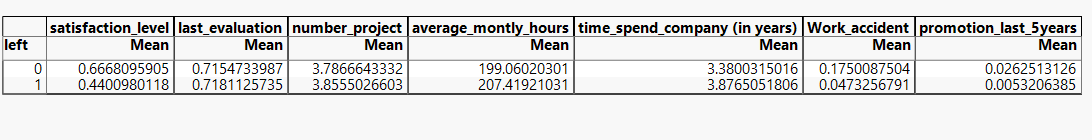


**Observations:**

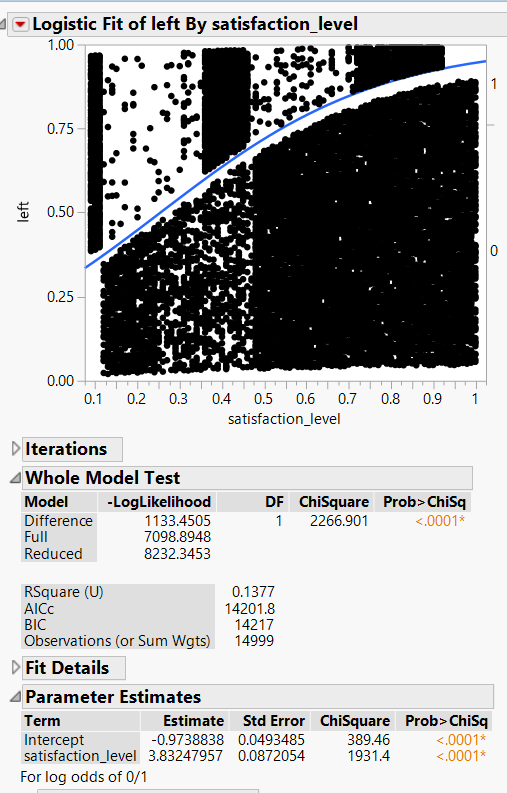
As per our analysis, we have observed that the data is clean except

* Time Spend Company: But as per the business requirements, we cannot exclude the outliers as it is necessary from our analysis perspective**.**

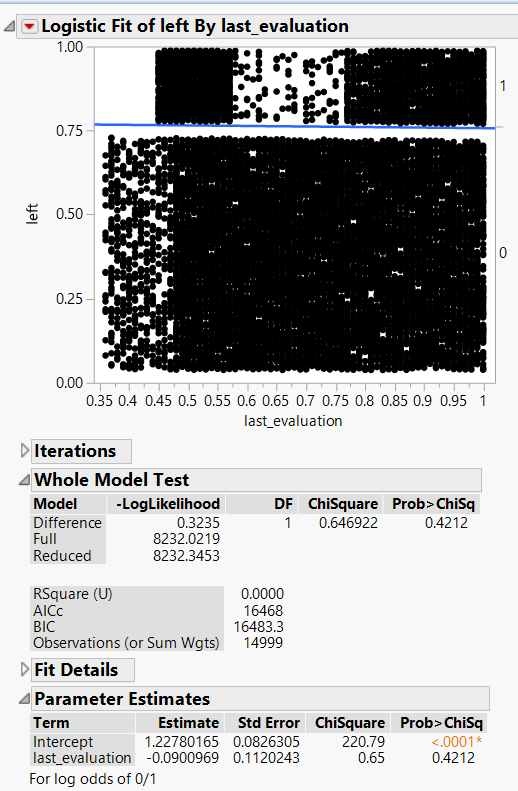
1. **Relationship among response and predictors**



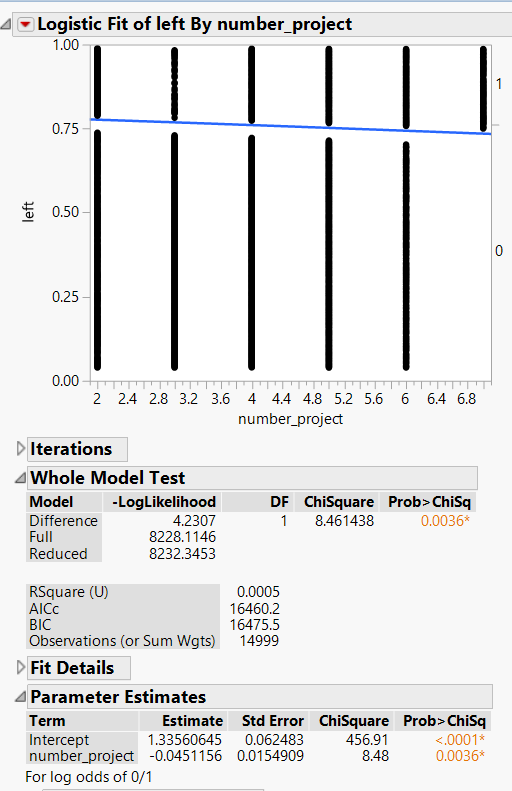
* Mean of the satisfied employees are 44% who left the company.
* Mean of unsatisfied people are 66%
* Average projects per employee who are the firm are 3.78
* Average projects per employee who left in firm are 3.85
* Average number of monthly hours for employees who are the firm are 199 hrs.
* Average number of monthly hours for employees who left the firm is 207 hrs.



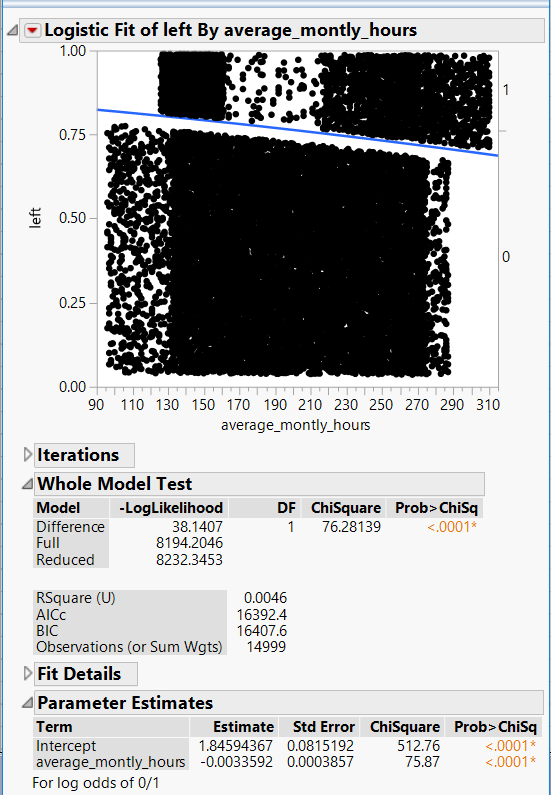
* From the above graph, we can see that there is no direct pattern in the data and no evident conclusions can be drawn about the people who left the firm or not.
* The BIC, AIC values are higher which says that there is not enough evidence to get any concrete results



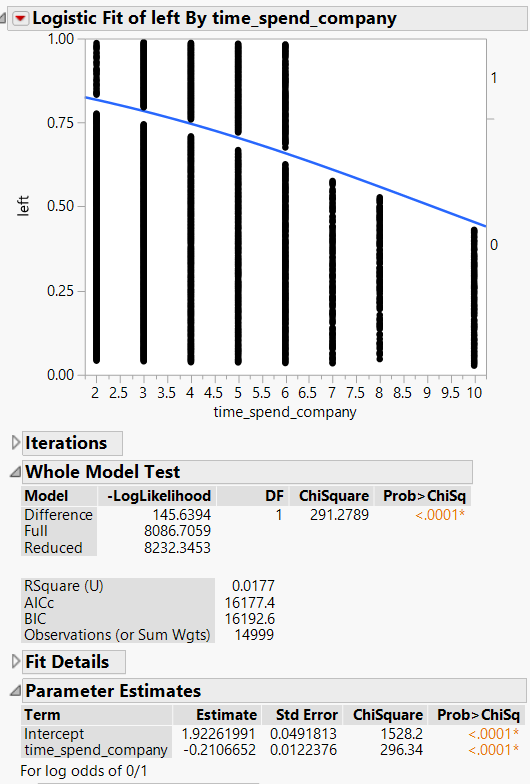
* From the above graph, we can see that there is no direct pattern in the data and no evident conclusions can be drawn about the people who left the firm or not.



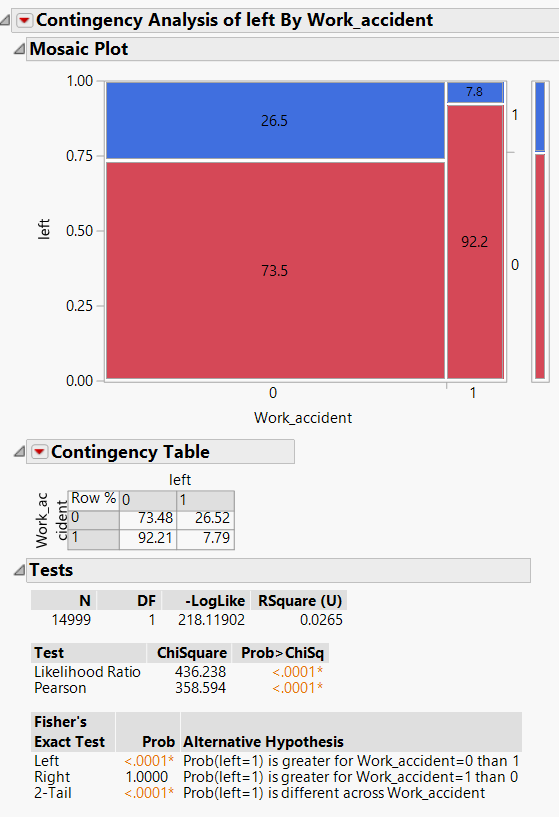
* From the above graph, we can see that there is no direct pattern in the data and no evident conclusions can be drawn about the people who left the firm or not.



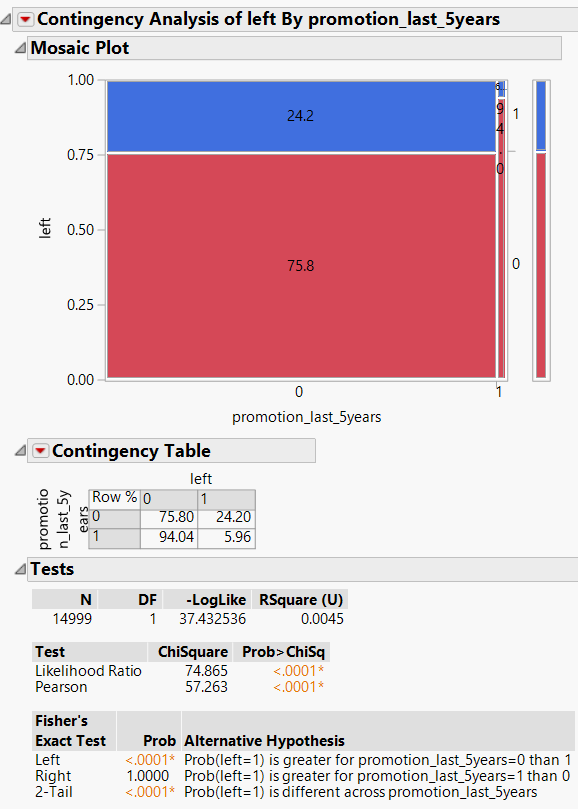
* From the above graph, we can see that there is no direct pattern in the data and no evident conclusions can be drawn about the people who left the firm or not.

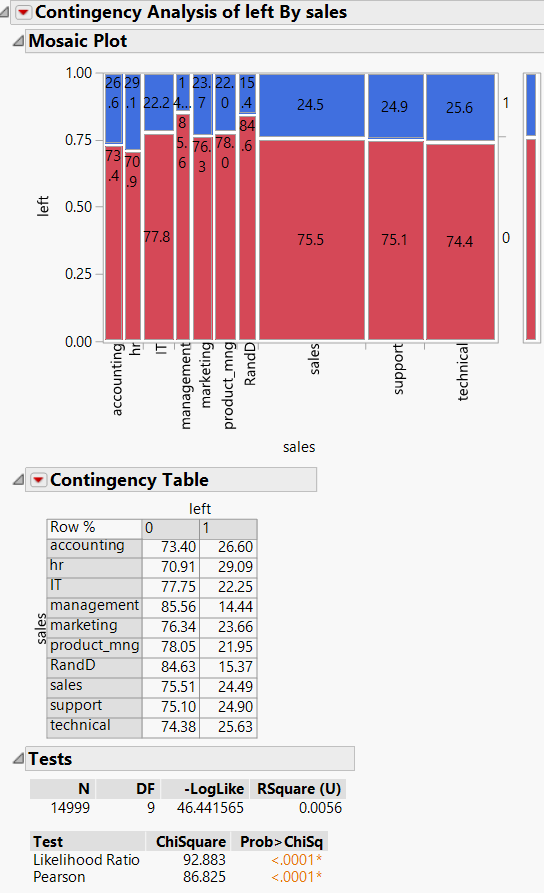


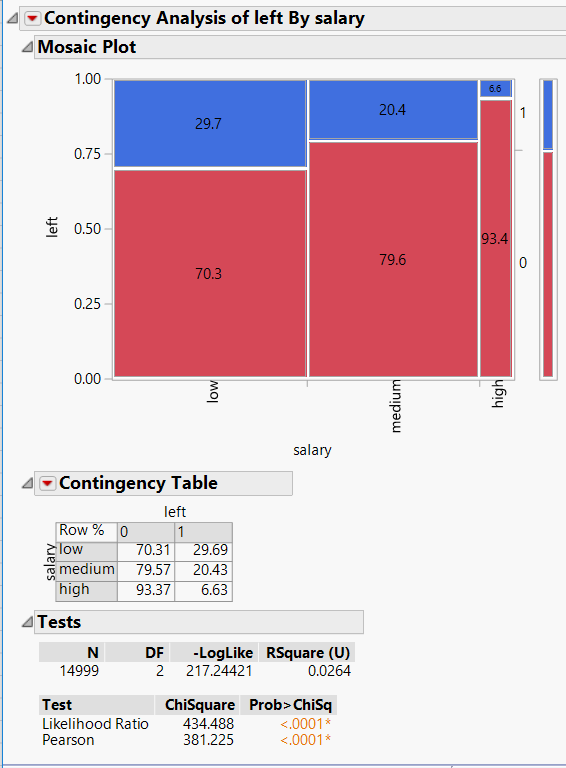
* **From the above graph, we can see that there is no direct pattern in the data and no evident conclusions can be drawn about the people who left the firm or not.**



* From the above graph, we can see that 92.2% of the people left because of work accident.

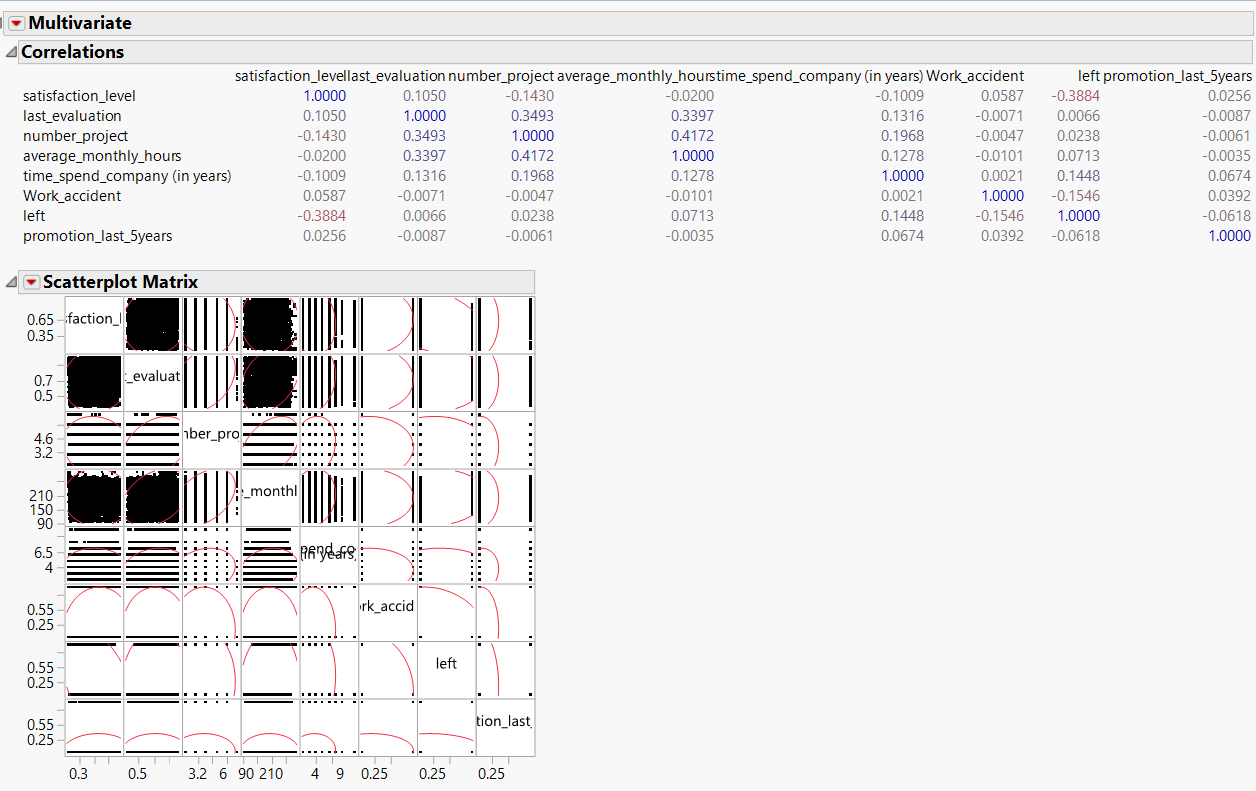


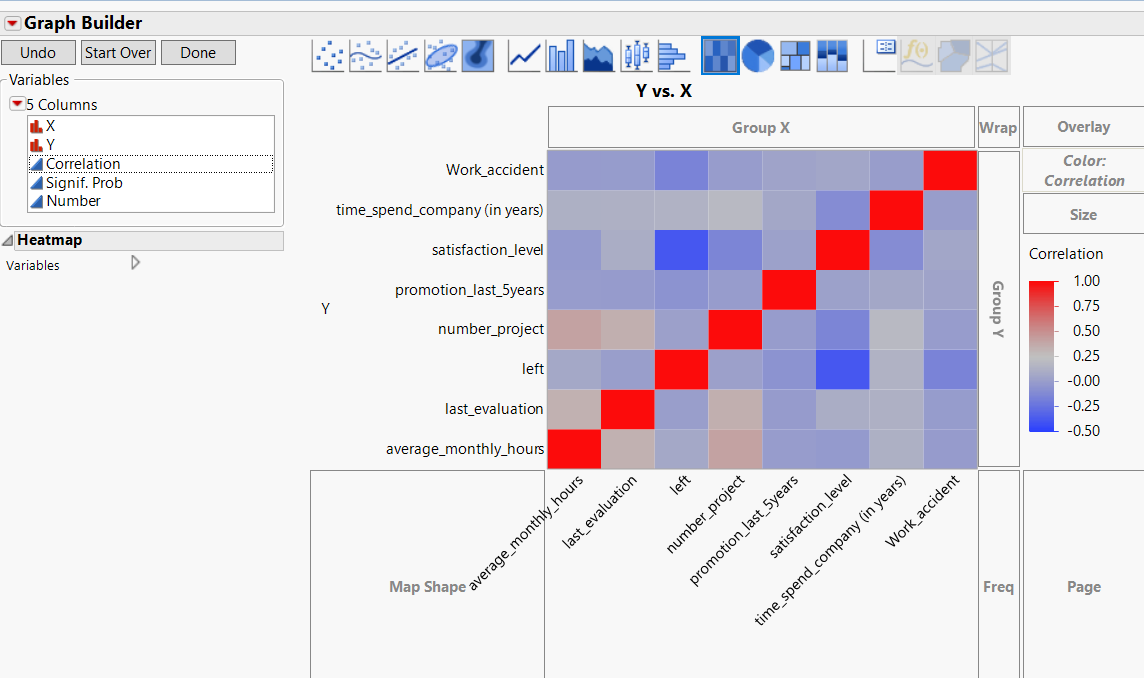




* Above mosaic plot shows the relation between salary grade and who left**.**

**2.2 Correlation between variables**



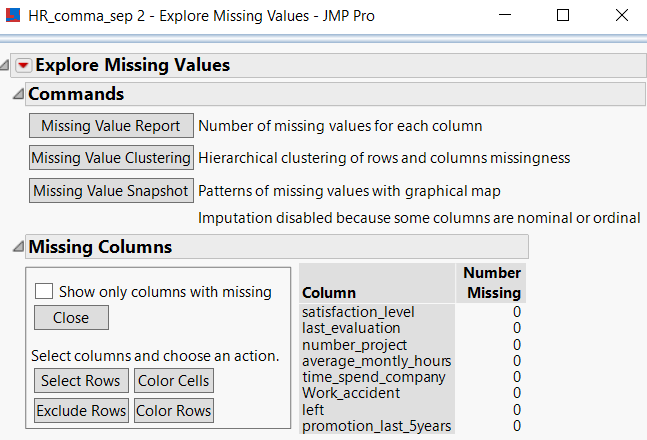


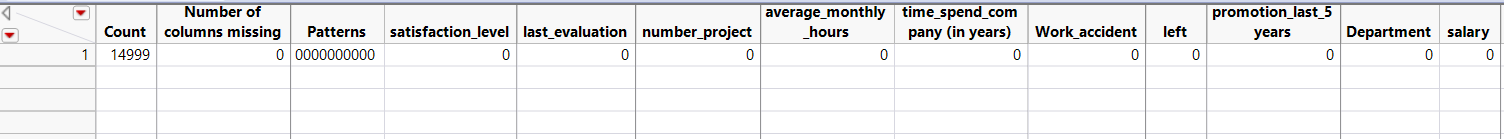
**For the correlation analysis, we plotted two graphs and the observation from them are as follows:**

* As per the graph, we can see there is no significant correlation between the target variable and the predictors except between the satisfaction level and the people who left.

**Data Modification**

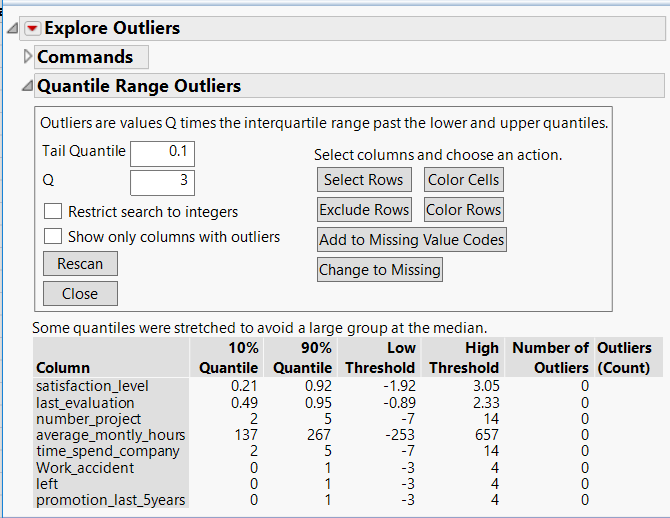
* 1. **Exploring Missing Values**





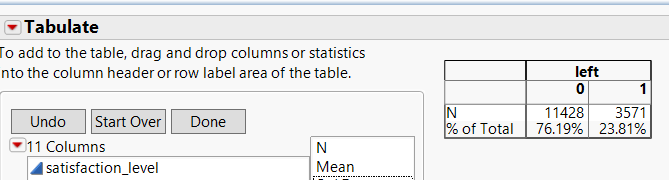
We performed the missing value analysis by exploring missing value pattern and we found that our data didn’t contain any missing value.

* 1. **Outlier Analysis**



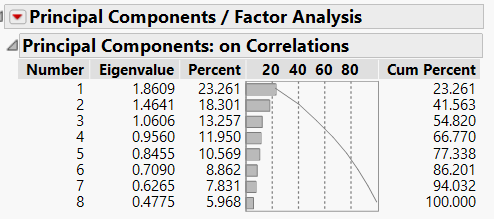
As per our business requirements and data analysis, we found that we don’t need to impute any outliers or missing values**.**

**Employee Turnover**



* 1. **Data Reduction**

**PCA**



All variables are required in the model, therefore Not combining variables for building model.

**Data Modeling**

* 1. **Model Building**

**Partitioning the Data set –**

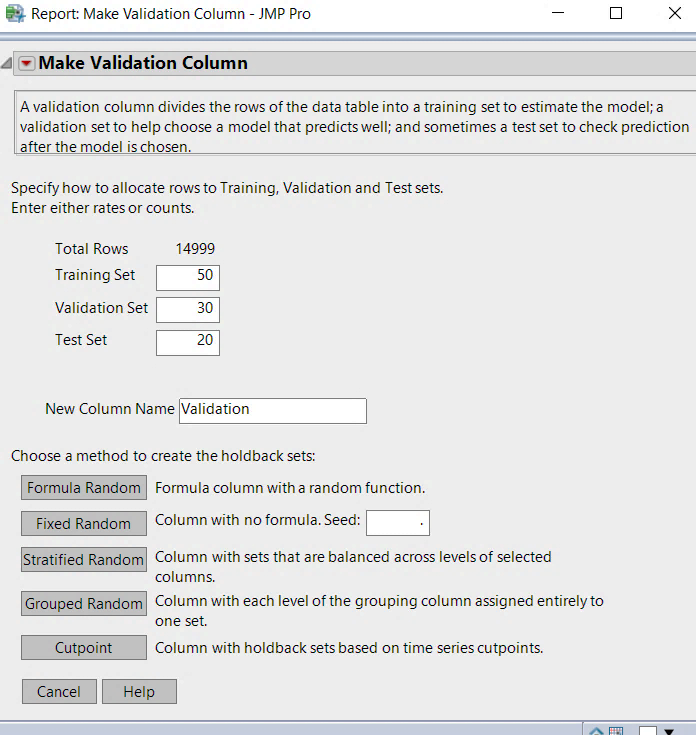
Data has been divided in to 3 sets as follow –

Training - 50%

Validation – 30%

Test – 20%

To avoid over fitting and to increase the accuracy, significant percentage of data has been divided to Validation data set.



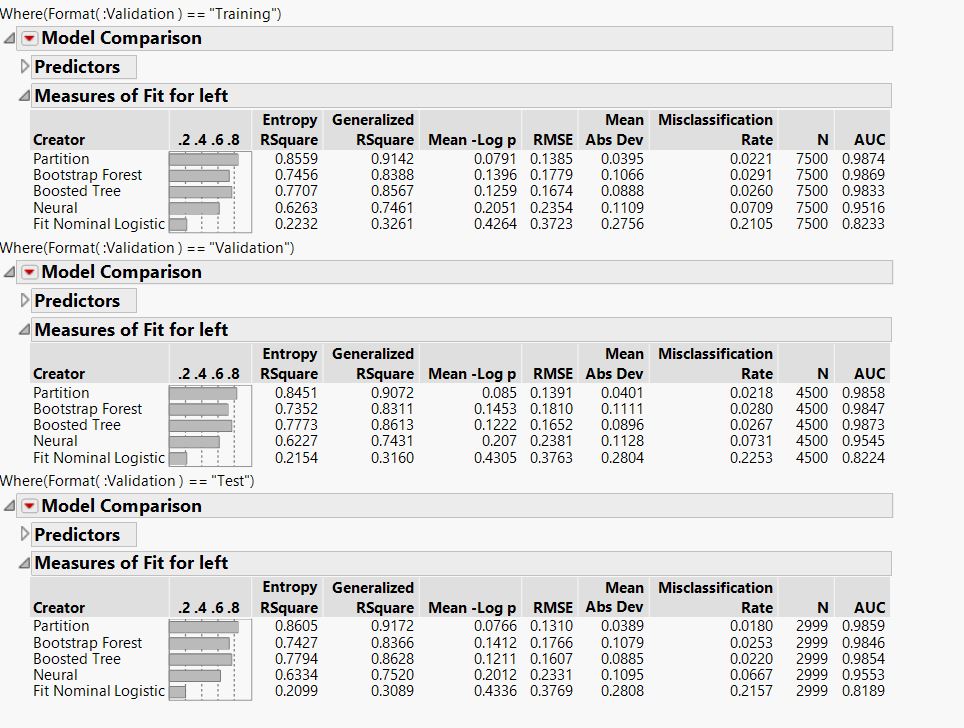
As our target variable is Nominal Variable, Data has been stratified random based on our target variable.

**Various Models used for our problem -**

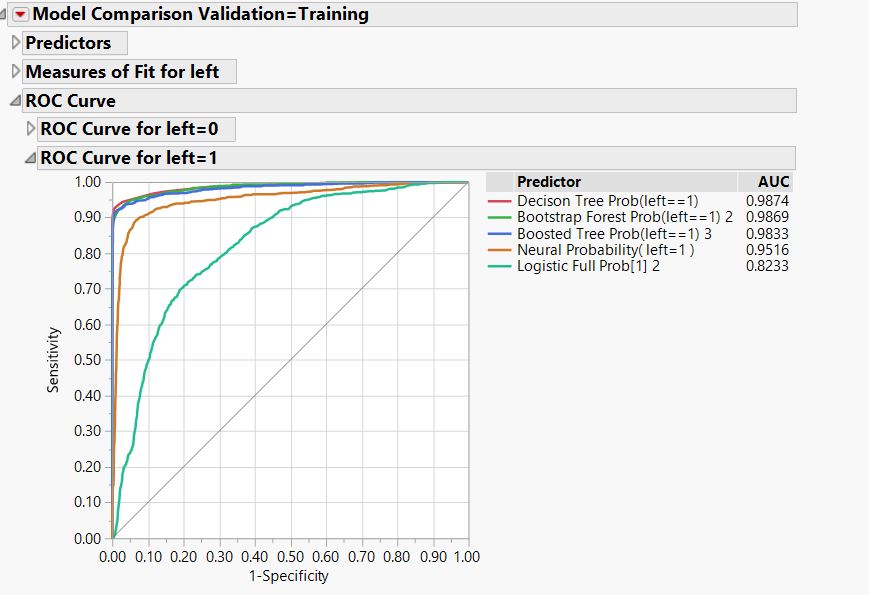
We have tried below classification models on our data set to find the best fit model and the model that predicts with best accuracy –

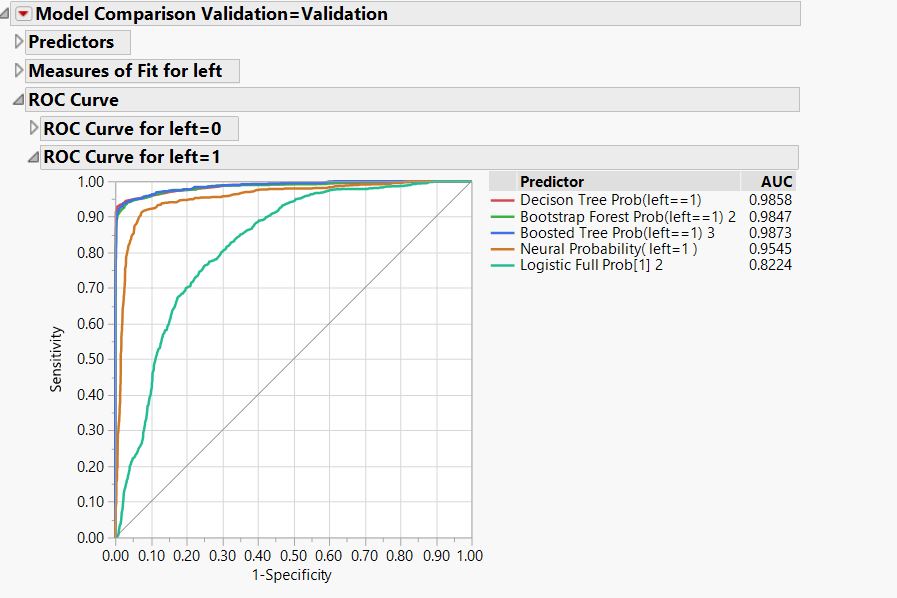
* Nominal Logistic
* Decision Tree (Partition)
* Boosted Tree
* Bootstrap Forest
* Neural

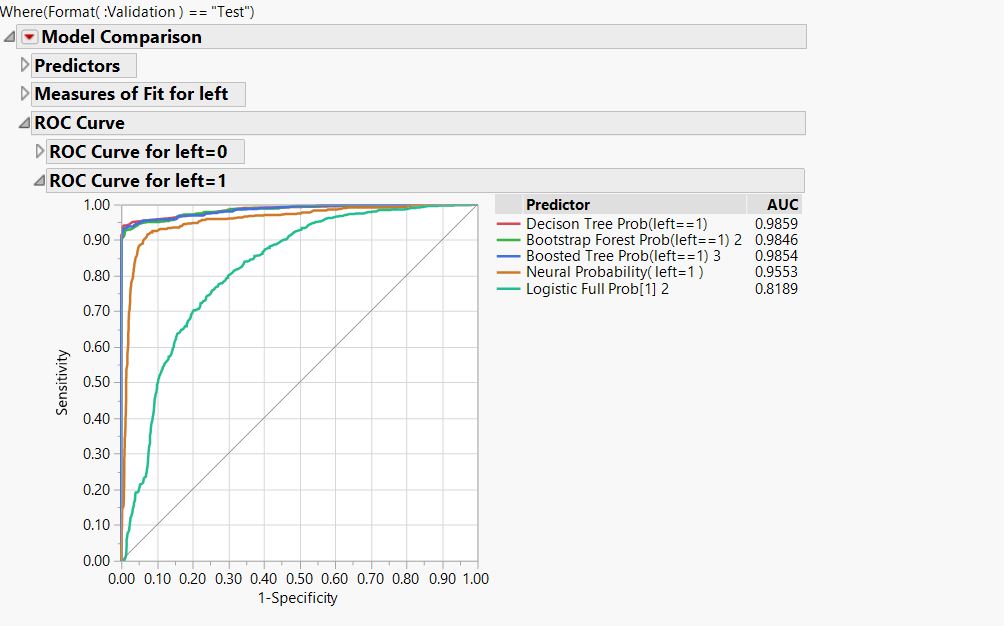
Data models have been built for all above algorithms and evaluated the performance of each model based on RMSE, Misclassification Rate, Generalized RSquare, AUC and Lift values. Model comparison was done to see which model has highest accuracy in prediction.



In Validation Data set We can notice Partition Model has Highest Generalized RSquare and, least RMSE and Misclassification Rate.

* **Choose the best fit model**

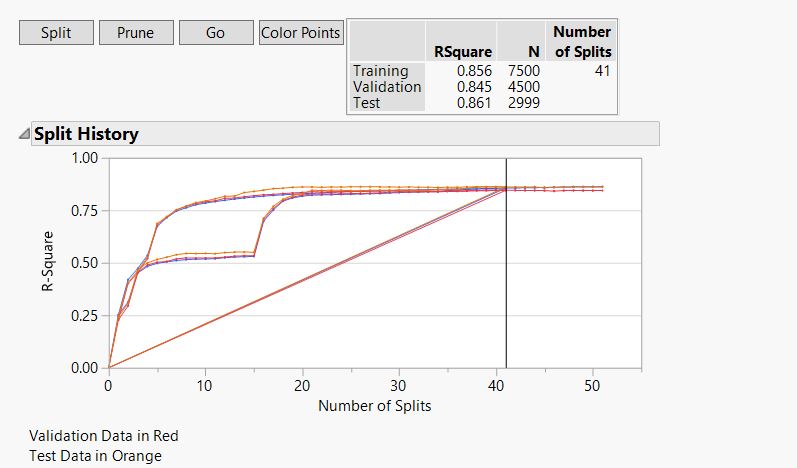




From the above ROC curves, we can notice that Partition model has the highest AUC. Interestingly all the performance evaluation parameters are best for Partition model in Training, Validation and Test Data sets.

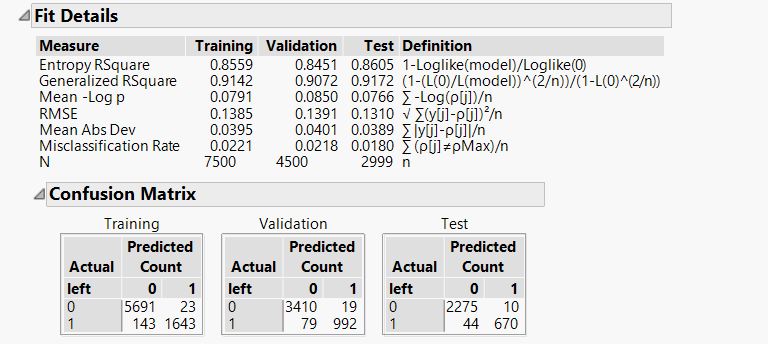
**More on Partition Model –**

**Split History –**

****

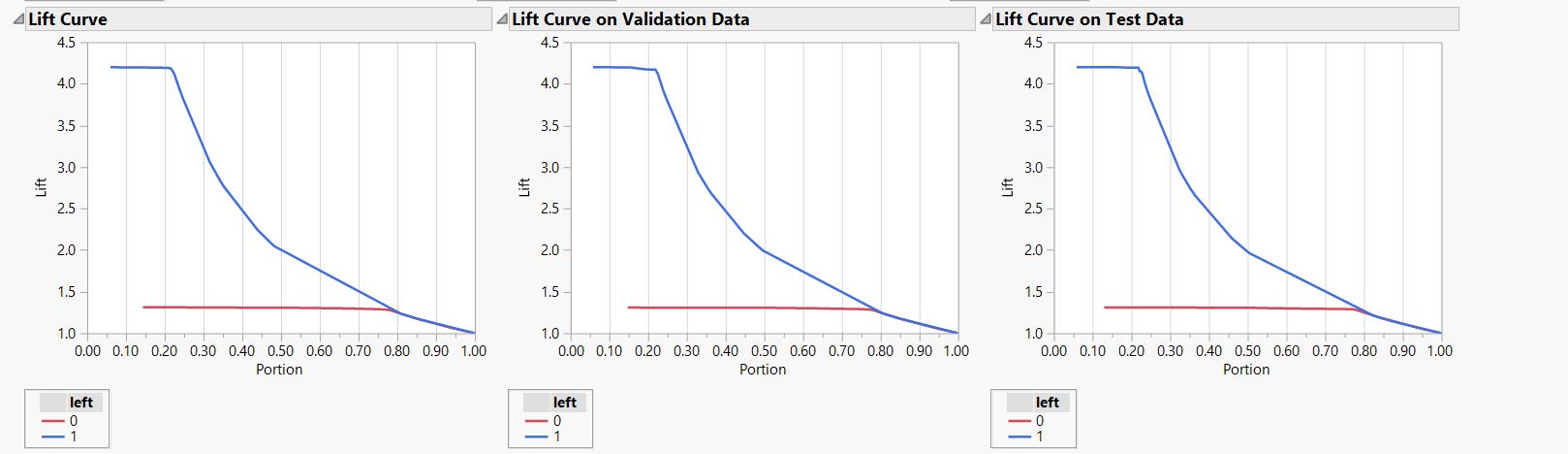
We can notice after 41 splits; R Square became constant.

**Fit Details –**

****

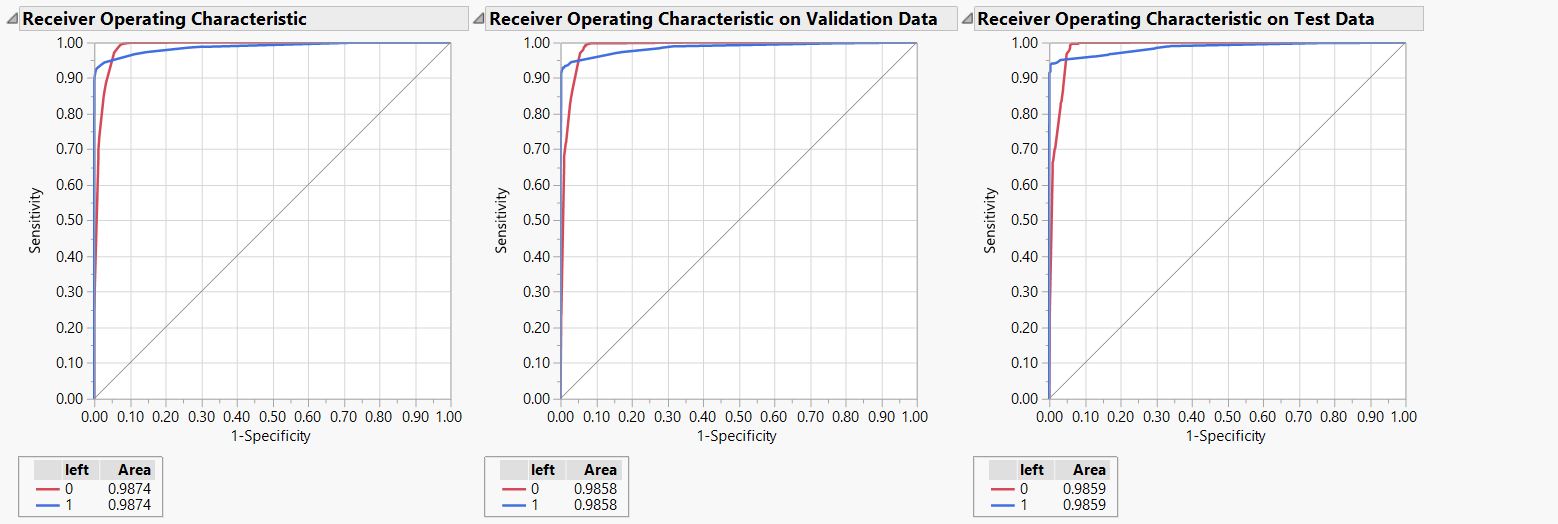
All the values are close by in Training, Validation and Test data set which implies that there is no Overfitting or under fitting in training Data set.

**Lift Curve –**

****

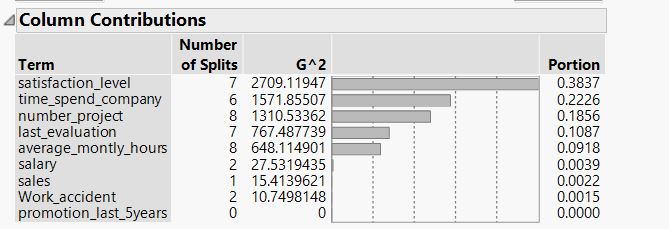
From the Lift curve, we can infer that 20% of data has lift of 4.25 implies that top 20% of data has 4 times more likely chance of leaving the company than the bottom 10% data. 80% of Data has lift of above 1.25.

**ROC Curve –**

****

The accuracy is 98.58% for validation data set. High value of AUC indicates that there is very low misclassification rate.

**Column contributions –**

****

Satisfaction Level is playing major role in predicating whether employee will leave the company or not followed by Time spend at company, Number of Projects. Last evaluation and Average Monthly Hours.

**Conclusion**

As per our understanding and our data analysis, we found that partition has the highest R square value and least RMSE Value.

**Summary:**

1. Employees left when they are working ~6hours/day and ~10hours/day
2. Employees that left had LOW-MEDIUM salary
3. Employees who left the firm did not get promoted
4. Employees who left had evaluation scores below 0.6 and above 0.8
5. Majority of employees who left had 2 projects. Every employee with 7 projects left.

**Recommendations**



As per our experience and understanding of the situation:

* Employee engagement can go a long way in reducing the attrition rate, as it helps in increasing the satisfaction level of the employee which intern helps in increasing the loyalty towards the firm which intern can transform into productive work which will intern will impact the revenue.
* Recognition can go a long way in increasing the confidence of the employee which intern will impact the revenue of the firm which will further decrease attrition rate
* Growth is a misconceived term, growth can be horizontal as well, which can motivate the employee to stay in the firm and hence reducing the attrition rate
* Root cause analysis: every employee has its own concern which can be resolve through one on one conversations and can help the firm in keeping their employees