Customer Churn Analysis

Using SAS Enterprise Miner and R Studio

By Sazee S. (Maninderpreet Singh Puri)

Table of Contents

1. Introduction

The case study is based on the use of churn analytics for Saturn Telecommunications to redesign the customer retention strategy. The aim is to understand the characteristics of churned, non-churned, and loyal customers. Further, to predict the customer propensity to churn, we are developing models. Also, we are recommending potential campaigns to buy back or win back the valued customers who are churned. The dataset provided *'telco_churn.csv'* has eighteen parameters and 7032 observations. It represents the existing customer base with their demographic's information, accounts information and service status recorded. The target variable in our study is *'Churn'* with values *'yes/no'*. To determine loyal non-churned customers, we are considering the longest *Tenure* of the customer's account opened with us in months. We are using SAS Enterprise Miner for Descriptive Analysis, Rstudio for calculating overall and group rates, and for developing/evaluating predictive models. The analysis is divided into three main tasks:

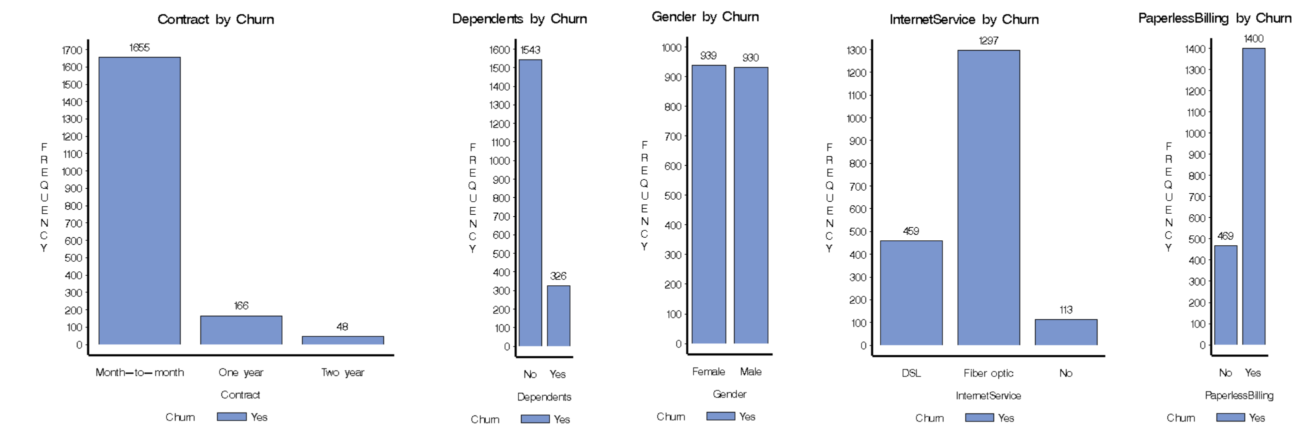
- To perform descriptive analysis on customer data and identify the characteristics of churned, non-churned and loyal customers.
- To calculate churn rates and develop three predictive models, compare their accuracies to find the best model.
- To provide campaign recommendations based on insights gained through our analysis.

The first step in a data analytics task is data pre-processing. The data used in our analysis consists of binary, nominal and interval variables and is free from missing values and skewness. So, we don't need any transformations for our variables. However, just to enhance our results we are applying standardization on interval variables before creating predictive models. Also, I set *'Churn'* to *Role* as *Target* and *'CustomerID'* as *rejected,* as it doesn't contribute to our study.

2. Understanding the characteristics of churned, non-churned customers and loyal customers

In this task I used SAS Enterprise Miner to perform descriptive analysis on our dataset. I used two *Filter* nodes to filter dataset with *Churn* as *Yes* and *No.* I used another *Filter* node for *Churn* as *No* dataset, to filter the top decile (top 10 percent) of non-churned customers based on the *tenure* variable. These are the loyal customers which are a subset of non-churned customers. I used *MultiPlot* and *Variable Clustering* nodes to create plots for analysis.

2.1 Churned Customers



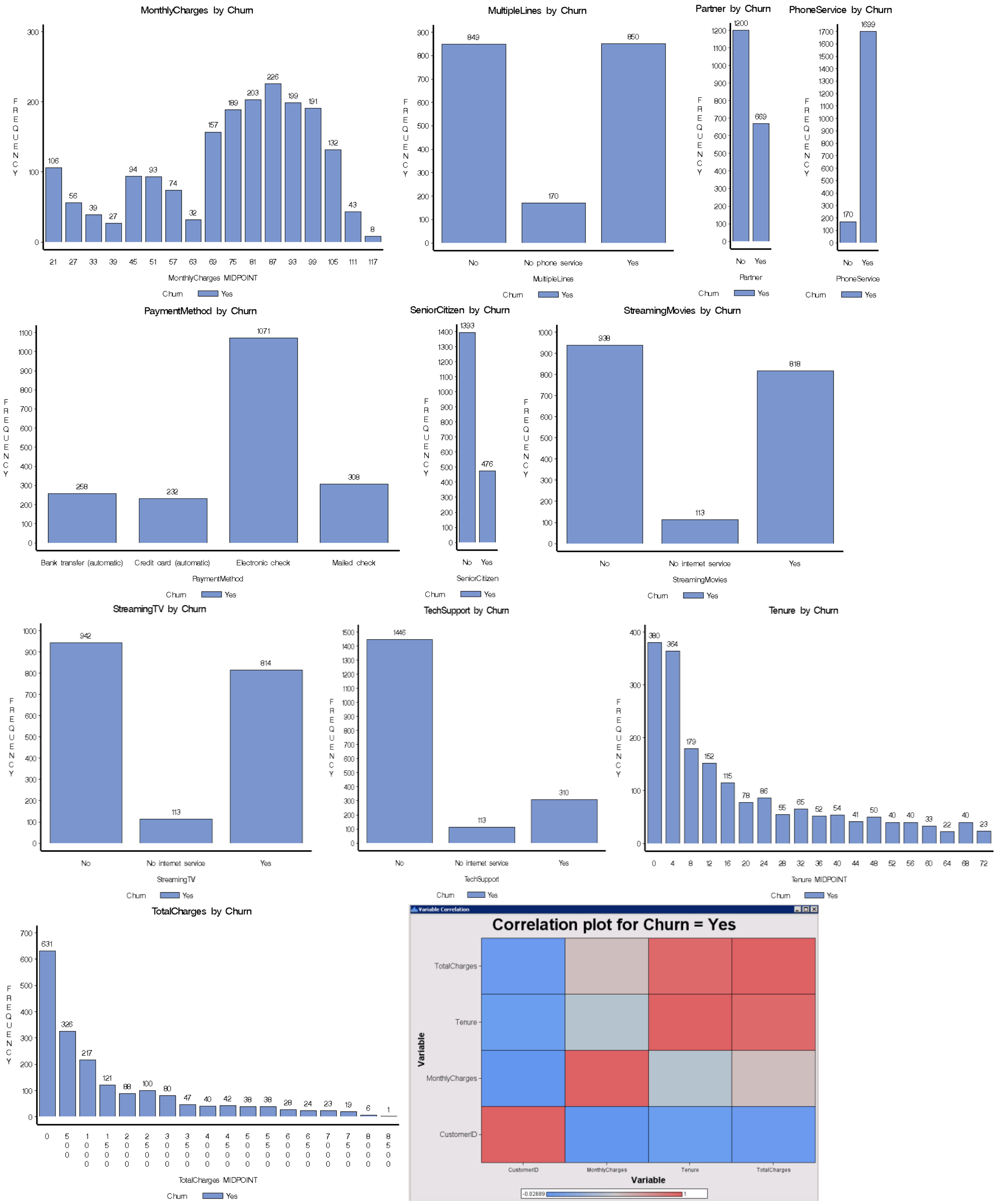
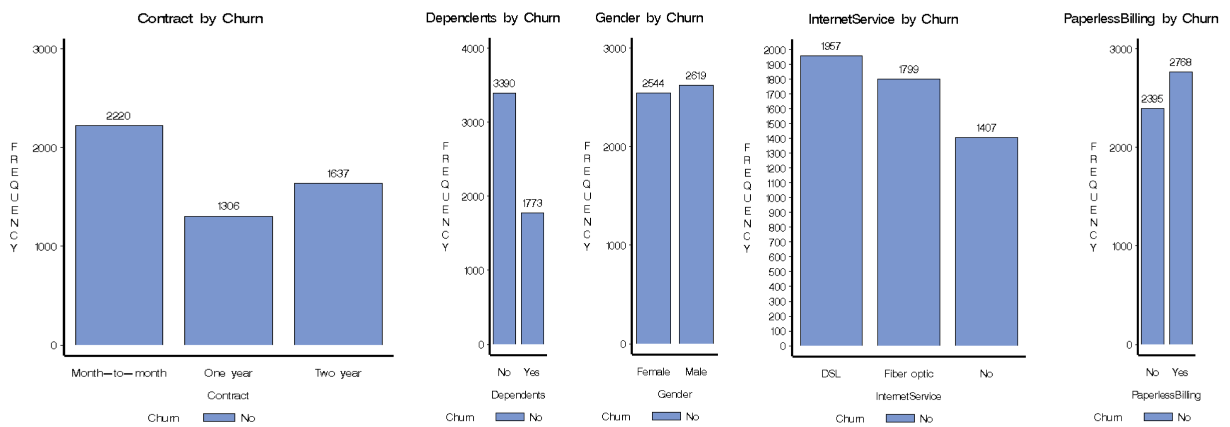


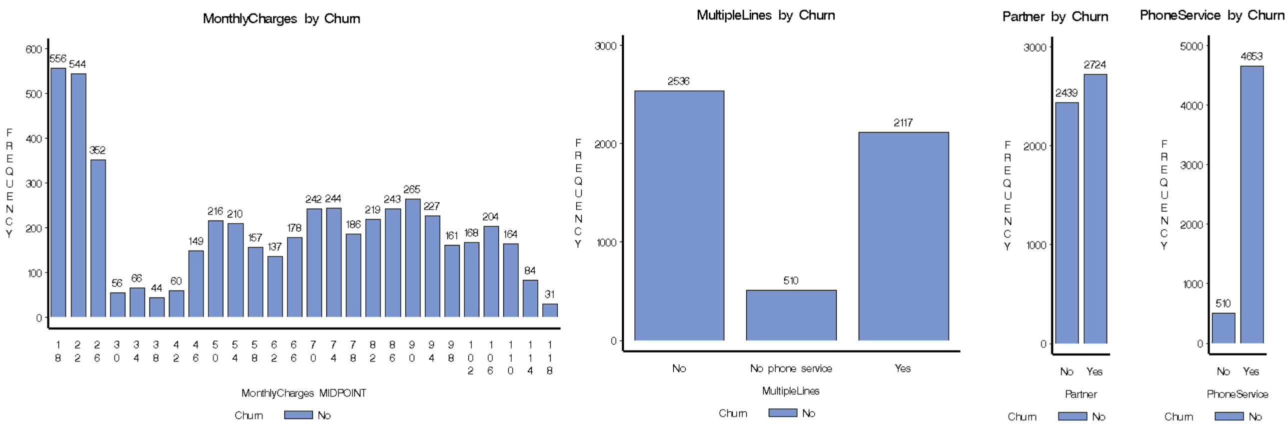
Figure 1 Plots for Churn = Yes data from Multiplot and Variable Clustering nodes.

In the plots above, we can see all the variables are plotted for *Churn = Yes* and dataset has *1869* observations. In the *Contract by Churn* plot we can see most of the customers have a *month by month contract (~89%),* and only a few customers have *one (~8.8%)* or *two year (~2.5%) contract*. *Dependents by Churn* plot tells most people don't have *Dependents (~82.5%)*. *Gender by Churn* plot tells number of *Males* and *Females* are almost same. *InternetService by Churn* tells most people have *Fiber Optic (~69%)*, followed by *DSL (~25%)*, and *No (~6%)*. In *PaperlessBilling by Churn* plot, we can see most people have opted for this service *(~75%)*, and rest have not. In *MonthlyCharges by Churn* plot shows most customer *(~69%)* have monthly charges between *69 to 105*. In *MultipleLines by Churn* plot shows *(~45%)* people say *yes* and *(~45%)* say *no to MultipleLines* and *(~9%)* opt for *no phone service*. In *Partner by Churn* plot, we can see that *(~64%)* don't have a partner. In *PhoneService by Churn* plot, we can see *(~91%)* say Yes to *PhoneService.* In *PaymentMethod by Churn*, we can see people in this group prefer *Electronic Check (~57%),* followed by *Mailed Check (~16.5%), Bank transfer (automatic) (~14%), and Credit Card (automatic) (~12.5%).* In *SeniorCitizen by Churn,* we can see *(~75%)* are not senior citizens. In *StreamingMovies by Churn* plot we can see *(~50%)* people said *No*, *(~44%)* said *Yes,* and rest opted for *No internet service (~6%).* In *StreamingTV by Churn*, we can see it has almost the same distribution as *StreamingMovies by Churn.* In *TechSupport by Churn* we can see *(~77%)* take *No* techsupport, *(~16.5%)* say *Yes* to *techsupport,* and *(~6%)* say *No Internet Service.* In *Tenure by Churn*, we can see most people are between *0-16 months (~64%)*. In the *TotalCharges by Churn* plot, we can see that most people *(~63%)* have *0-1000 Totalcharges* so far and very few have a high total charge.

The correlation of the interval variables show a very high correlation *(0.95)* between *TotalCharges* and *Tenure* which makes sense as the charge increases with time. As we can infer from the information, most people in this group have a month by month contract, fiber optics internet service, paperless billing, monthly charge between 69-105, don't have a partner, pay via electonic check, don't opt for streaming movies and t.v, don't take tech-support, stay for an year, and pay very less for the services.

2.2 Non- Churned Customers





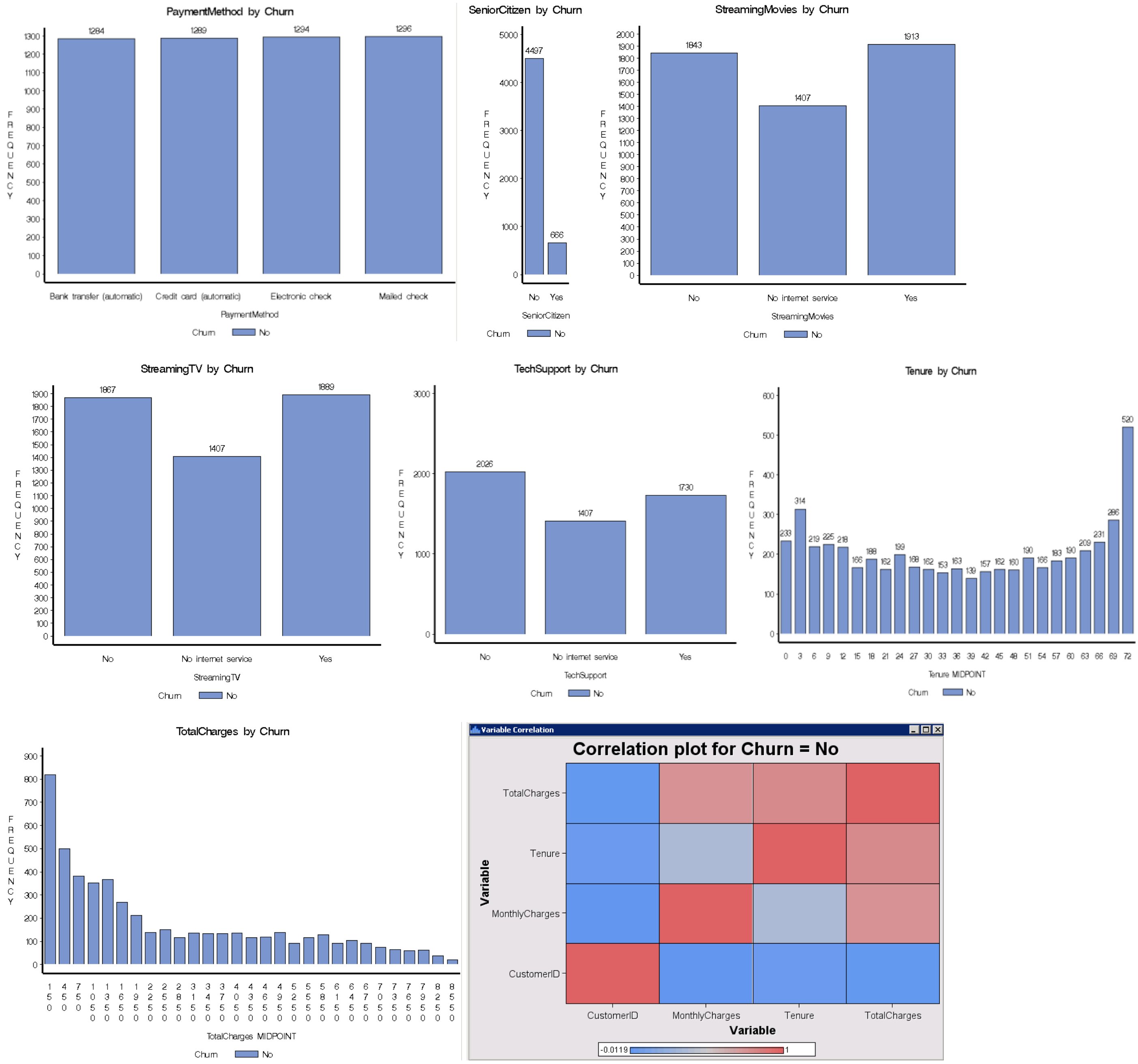


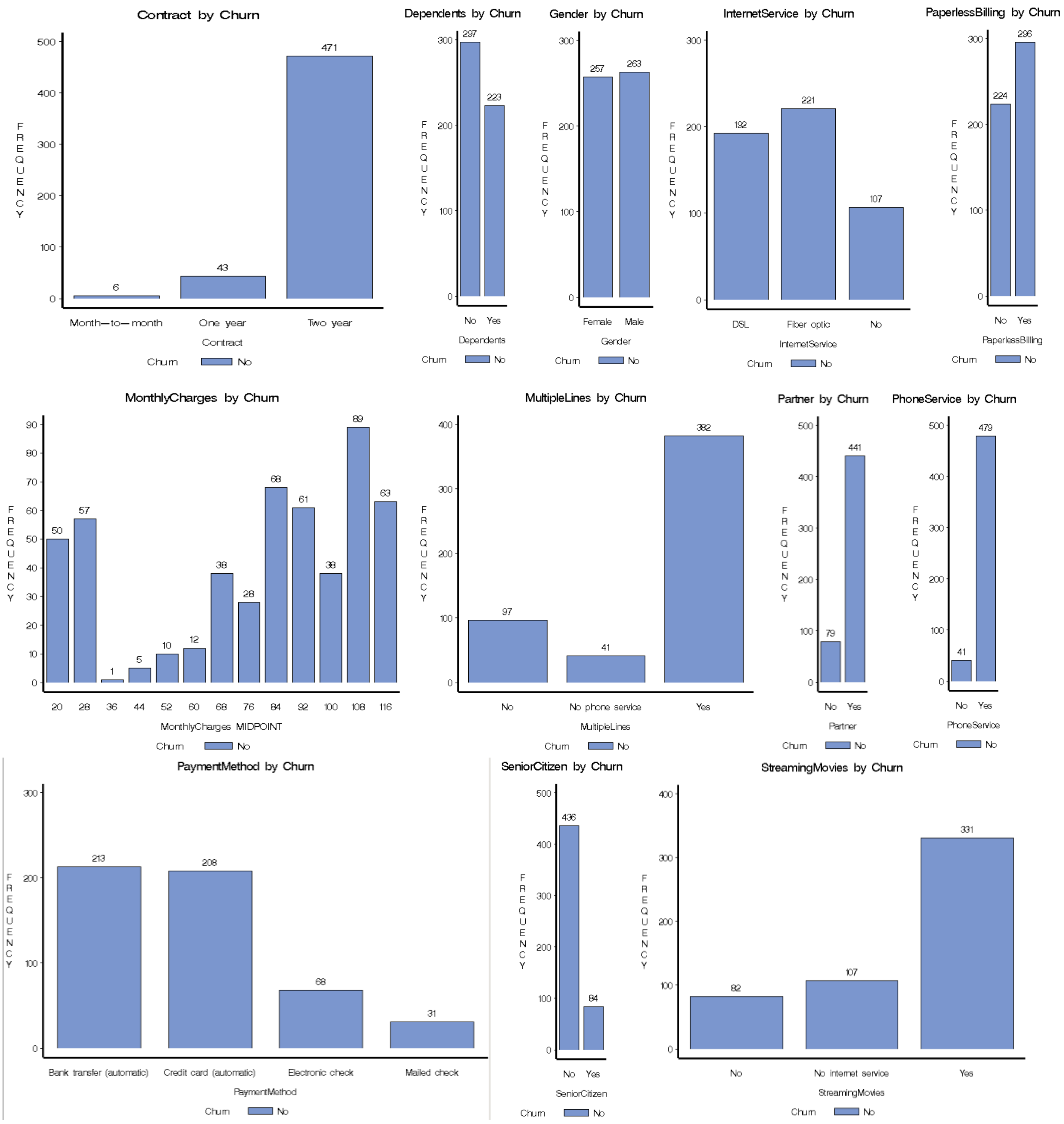
Figure 2 Plots for Churn = No data from Multiplot and Variable Clustering nodes.

In the plots above, we can see all the variables are plotted for *Churn = No* and dataset has *5163* observations. In the *Contract by Churn* plot we can see most of the customers have a *month by month contract (~43%),* some have a *two year (~32%) contract or one (~25%) contract*. *Dependents by Churn* plot tells most people don't have *Dependents (~66%)*. *Gender by Churn* plot tells number of *Males* and *Females* are almost same. *InternetService by Churn* tells most people have *DSL (~38%)*, followed by *Fiber optic (~35%)*, and *No (~27%)*. In *PaperlessBilling by Churn* plot, we can see almost half the population have opted for this service *(~54%)* and rest have not. In *MonthlyCharges by Churn* plot shows most customer *(~68%)* have monthly charges between *46 - 110* and others are between *18-26 (~28%).* In *MultipleLines by Churn* plot shows *(~49%)* people say *No* and *(~41%)* say *Yes* to *MultipleLines* and *(~9%) No phone service*. In *Partner by Churn* plot, we can see that *(~53%)* have a partner. In *PhoneService by Churn* plot, we can see *(~90%)* say *Yes* to *PhoneService.* In PaymentMethod by Churn, we can see people in this group equally prefer all payment methods *Electronic Check, Mailed Check, Bank transfer (automatic), and Credit Card (automatic).* In *SeniorCitizen by Churn,* we can see *(~87%)* are not senior citizens. In *StreamingMovies by Churn* plot we can see *(~37%)* people said *Yes*, *(~36%)* said *No,* and rest opted for *No internet service (~27%).* In *StreamingTV by Churn*, we can see it has almost the same distribution as *StreamingMovies by Churn.* In *TechSupport by Churn* we can see

(~39%) take *No* techsupport, *(~34%)* say *Yes* to *techsupport,* and *(~27%)* say *No Internet Service.* In *Tenure by Churn*, we can see *(~24%)* people are between *63-72 months, (~19%)* are between *0-9,* and rest are between *9-63 (~57%)*. In the *TotalCharges by Churn* plot, we can see that most people *(~53%)* have *0-1950 Totalcharges* so far and very few have a high total charge.

The correlation plot of the interval variables show a very high correlation between *TotalCharges* and *Tenure (0.79)* and *MonthlyCharges* and *TotalCharges (0.75)*. This makes sense as the total charge increases with time and monthly charge. As we can infer from the information, most people in this group have a month by month and two year contract, DSL internet service, paperless billing, monthly charge between 46-110, have a partner, opt for streaming movies and t.v, don't take tech-support, stay between an year and a half, and mostly pay less for the services.

2.3 Loyal Customers



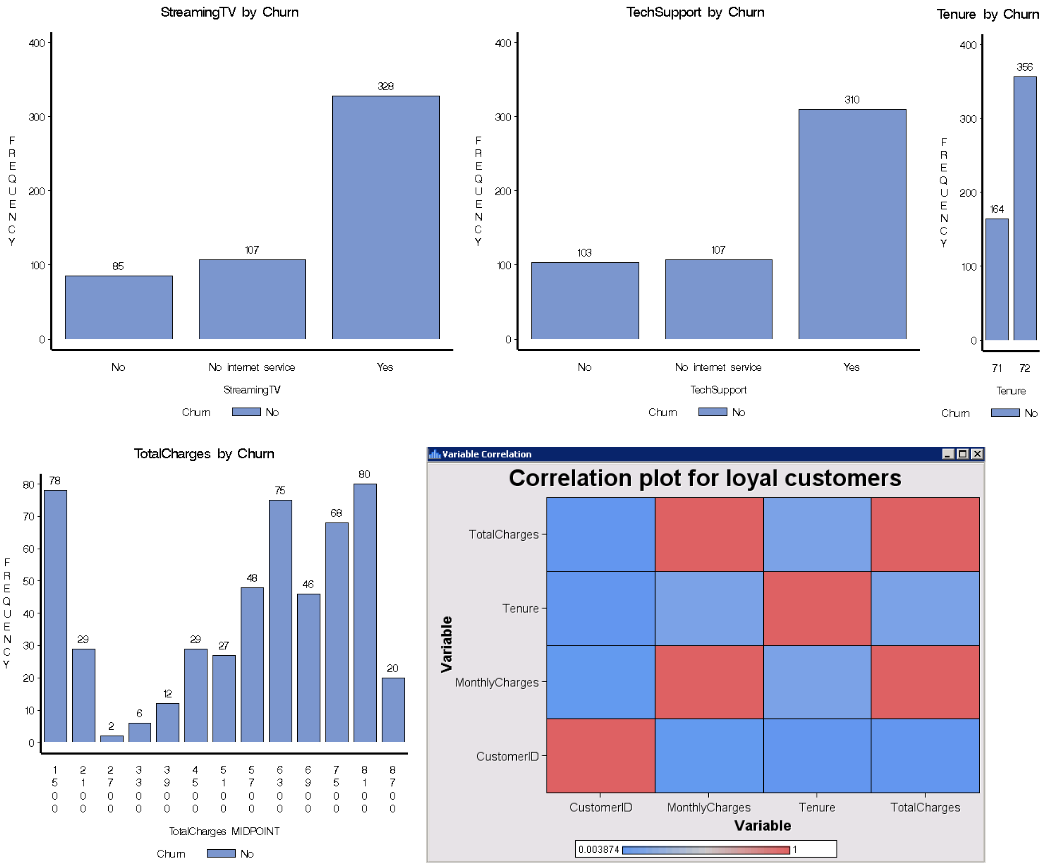


Figure 3 Plots for Churn = No and top decile based on Tenure data from Multiplot and Variable Clustering nodes.

In the plots above, we can see all the variables are plotted for loyal customers (*Churn = No and top decile)* and dataset has *520* observations*.* In the *Contract by Churn* plot we can see most of the customers have a a *two year (~91%) contract , one (~8%) contract,* and *month by month contract (~1%). Dependents by Churn* plot tells most people don't have *Dependents (~57%). Gender by Churn* plot tells number of *Males* and *Females* are almost same. *InternetService by Churn* tells most people have *Fiber optic (~43%)*, followed by *DSL (~38%)*, and *No (~21%)*. In *PaperlessBilling by Churn* plot, we can see almost half the population have opted for this service *(~57%)* and rest have not*.* In *MonthlyCharges by Churn* plot shows most customer *(~74%)* have monthly charges between *68-116* and others are between *20-68 (~26%)*. In *MultipleLines by Churn* plot shows *(~74%)* people say *Yes* and *No (~19%) to MultipleLines* and *(~9%) No phone service*. In *Partner by Churn* plot, we can see that *(~85%)* have a partner. In *PhoneService by Churn* plot, we can see *(~92%)* say *Yes* to *PhoneService.* In *PaymentMethod by Churn*, we can see people in this group equally prefer *Bank transfer (automatic), and Credit Card (automatic).* Less people prefer *Electronic Check and Mailed Check* payment methods*.* In *SeniorCitizen by Churn,* we can see *(~84%)* are not senior citizens. In *StreamingMovies by Churn* plot we can see *(~64%)* people said *Yes*, *(~21%)* said *No,* and rest opted for *No internet service (~16%).* In *StreamingTV by Churn*, we can see it has almost the same distribution as *StreamingMovies by Churn.* In *TechSupport by Churn* we can see *(~60%)* say *yes* to *techsupport*, *(~21%)* say *no* to *techsupport,* and *(~20%)* say *No Internet Service.* In *Tenure by Churn*, we can see *(~69%)* people have *72 months,* and *(~31%)* have *71 months* long tenure. In the *TotalCharges by Churn* plot, we can see that most people *(~65%)* have *5700-8700 Totalcharges* so far and a few have a total charge between *1500-5700 (~35%)*.

The correlation plot of the interval variables show a very high correlation between *TotalCharges* and *MonthlyCharges (0.99)*. This makes sense as the total charge increases with monthly charge. As we can infer from the information, most people in this group have a two year contract, fiber optic internet service, paperless billing, monthly charge between 68-116, have a partner, pay via bank transfer or credit card (automatics), opt for streaming movies and t.v, take tech-support, stay more than five years, and mostly pay high for the services.

We can see all the customer profiles are very unique and have different attributes. Now we can use this data to perform futher analysis and provide recommendations.

3. Developing and evaluating models to predict propensity to churn

In this part of analysis, we are calculating churn rates for overall dataset and categorical variables. Then we are developing prediction models to predict propensity to churn.

3.1 Overall and Group Churn Rates

I used RStudio to get overall and group churn rates. The results from Rstudio show *Overall Churn rate* is: 0.*27*. This means almost one-fourth of the customers churn in a normal situation. For group churn rates I calculated values for all categories group wise. The results show group churn for *PaymentMethod (Electronic check)* is the highest 0.*45.* This means that people who use *PaymentMethod (Electronic)* are more likely to churn as compared to all other customers.

3.2 Developing prediction models

I have used three predictive models: Regression, Decision Tree, and Neural Network. I have used *Transform Variables* node and set all interval variables *(MonthlyCharges, TotalCharges, Tenure) Method* to *Standardize* in *Edit variables* option. Also, I have used the *Data Partition* node to partition the data into training (70%) and validation (30%) data sets. Further, I used *Variable clustering* (with *Variable Selection* as *Best Variables*) to reduce the number of similar variables (in terms of correlation) for our modelling. Finally, I used *Model Comparison* node to compare all models and get the best model.

Sample Statistics

Obs #	Variable Name ▼	Label	Type	Percent ...	Minimum	Maximum
4	TotalCharges		VAR	0	18.8	8684.8
3	Tenure		VAR	0	1	72
2	MonthlyCharges		VAR	0	18.25	118.75

Figure 4 Range for TotalCharges, Tenure and MonthlyCharges variables.

3.2.1 Reasons for Transforming Interval variables

We are applying the standardization on interval variables as these variables have very different range of values. Predictive models need variables with same range of variables to make correct predictions. For example, in *MonthlyCharges* values range from *18.25 to 118.75*, whereas *TotalCharges* ranges from *18.8 to 8684.8*. The *Tenure* variable has the lowest range from *1 to 72*. So, when we apply standardization, the values of all variables are squished in a range which is common across all the variables, so that when developing the prediction model all variables are given equal weightage.

3.2.2 Selected variables used for building the prediction models

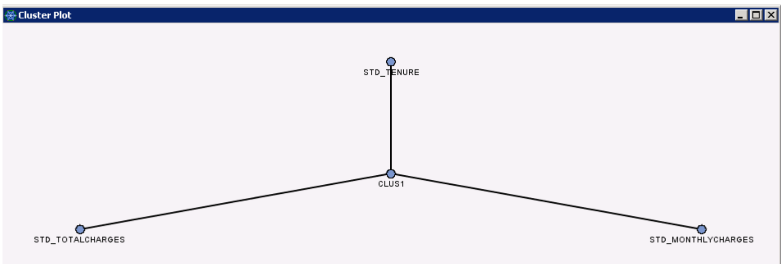


Figure 5 Cluster plot for TotalCharges, Tenure and MonthlyCharges variables.

Variable Selection Table

Cluster	Variable	Label	R-Square With Own Cluster Component	Next Closest Cluster	R-Square with Next Cluster Component	Type	1-R2 Ratio	Variable Selected
CLUS1	CLUS1	Cluster 1	1	CLUS1	0	ClusterComp	0	NO
CLUS1	STD_TOTALCHA...	Transformed Tot...	0.96486	CLUS1	0	Variable	0.03514	YES
CLUS1	STD_TENURE	Transformed Te...	0.699533	CLUS1	0	Variable	0.300467	NO
CLUS1	STD_MONTHLY...	Transformed Mo...	0.511318	CLUS1	0	Variable	0.488682	NO

Figure 6 Variable Selection Table from Variable Selection node showing selected variable TotalCharges.

We have selected fourteen variables for building our predictive models. In our prediction models we are using *Churn, Contract, Dependents, Gender, InternetService, MultiplLines, PaperlessBilling, Partner, PaymentMethod, PhoneService, STD_TotalCharges, SeniorCitizen, StreamingMovies, StreamingTV, and TechSupport* variables. We have dropped two variables *Tenure* and *MonthlyCharges* from our analysis. In the *Variable Clustering* node, we can see in the *Cluster plot*, and the *Variable Selection Table*; that the *TotalCharges, MonthlyCharges,* and *Tenure* share similarity (high correlation). So, these variables are represented by a single variable *TotalCharges* in our subsequent analysis.

3.2.3 Predictive performance of various models

We are considering the validation results for our study to evaluate the models as training results will not give us actual performance of our model.

1. Regression

In the regression model, I have set the *Regression Type* to *Logistics Regression* as we are classifying our target variable. After running the model, I see *Score Rankings Overlay: Churn, Fit Statistics, Output,* and *Effects Plot* in the results window.

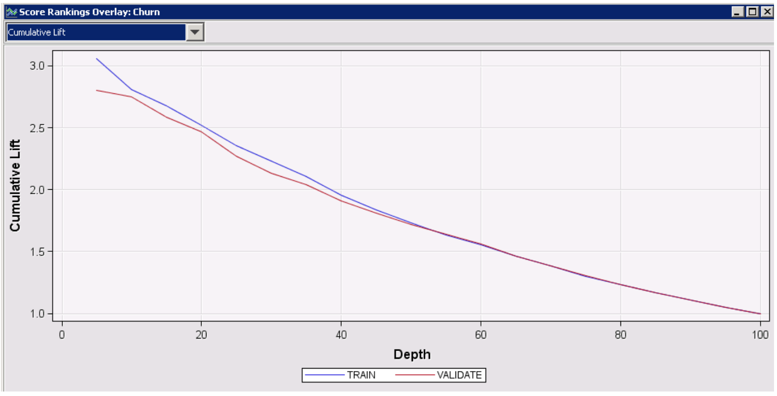


Figure 7 Score Rankings Overlay: Churn plot showing cumulative lift vs depth.

The *Score Rankings Overlay: Churn* show the plot for *Cumulative Lift* vs *Depth* of the model. It shows that *Training* phase has a slightly higher *Cumulative Lift* than *Validation* phase. The *Cumulative Lift* of around *three* means that this model can approximately predict three times better than a random guess for 20% of the customers (till Depth = 20). If we increase the number of observations the cumulative lift drops and will edge off to value one when we take all observations (Depth = 100) into consideration. This means when we consider all customers there is no difference in the model prediction and a random guess.

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation
Churn		_AIC_	Akaike's Information C...	4173.431	.
Churn		_ASE_	Average Squared Error	0.136455	0.141715
Churn		_AVERR_	Average Error Function	0.420267	0.432859
Churn		_DFE_	Degrees of Freedom f...	4901	.
Churn		_DFM_	Model Degrees of Free...	19	.
Churn		_DFT_	Total Degrees of Free...	4920	.
Churn		_DIV_	Divisor for ASE	9840	4224
Churn		_ERR_	Error Function	4135.431	1828.397
Churn		_FPE_	Final Prediction Error	0.137513	.
Churn		_MAX_	Maximum Absolute Error	0.98812	0.989548
Churn		_MSE_	Mean Square Error	0.136984	0.141715
Churn		_NOBS_	Sum of Frequencies	4920	2112
Churn		_NW_	Number of Estimate W...	19	.
Churn		_RASE_	Root Average Sum of ...	0.369399	0.376451
Churn		_RFPE_	Root Final Prediction E...	0.370828	.
Churn		_RMSE_	Root Mean Squared Er...	0.370114	0.376451
Churn		_SBC_	Schwarz's Bayesian Cr...	4296.952	.
Churn		_SSE_	Sum of Squared Errors	1342.721	598.6056
Churn		_SUMW_	Sum of Case Weights ...	9840	4224
Churn		_MISC_	Misclassification Rate	0.197358	0.206439

Figure 8 Fit Statics table showing statistic measures.

The *Fit Statics* window show the main statistics to determine the performance of the regression model. We can see that the *Misclassification Rate* for *Validation* dataset is *0.206439,* the *Average Squared Error* is *0.141715,* and the *Mean Squared Error is 0.141715.*

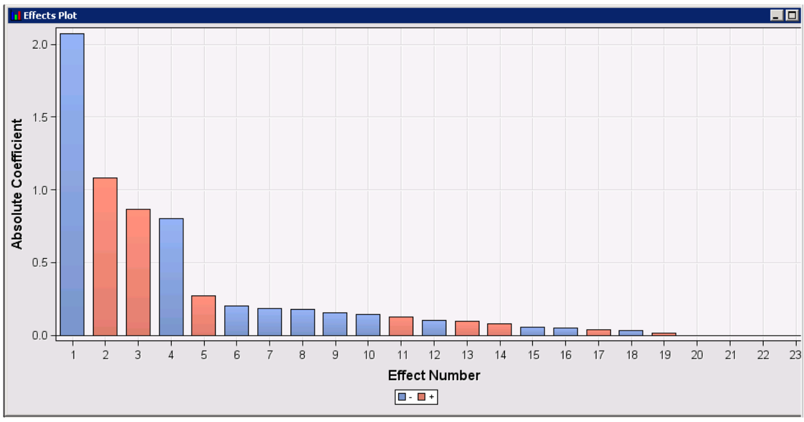


Figure 9 Effects plot showing Effects vs Absolute Coefficient.

The *Effects Plot show* the *Effects* in our linear model vs *Absolute Coefficient*. The first bar is the intercept which tells normally a customer would be not churned. The second bar represents *InternetServiceFIBER*_OPTIC which has an *Absolute Coefficient* of *1.078*. The red bar for churned customers. So, this means that if a customer opts for *InternetServiceFIBER_OPTIC*, the customer is more likely to churn than other customers. The results in this plot confirm our results obtained in Task 2-part a of our analysis as *InternetServiceFIBER_OPTIC, ContractMONTH_TO_MONTH, PaymentELECTRONIC_CHECK* (red bars) are the groups with the highest group churn rates.

Odds Ratio Estimates

Effect		Point Estimate
Contract	Month-to-month vs Two year	5.360
Contract	One year vs Two year	2.145
Dependents	No vs Yes	1.209
Gender	Female vs Male	1.025
InternetService	DSL vs No	2.640
InternetService	Fiber optic vs No	8.197
MultipleLines	No vs Yes	0.760
MultipleLines	No phone service vs Yes	0.983
PaperlessBilling	No vs Yes	0.738
Partner	No vs Yes	1.076
PaymentMethod	Bank transfer (automatic) vs Mailed check	1.020
PaymentMethod	Credit card (automatic) vs Mailed check	0.880
PaymentMethod	Electronic check vs Mailed check	1.381
PhoneService	No vs Yes	.
STD_TotalCharges		0.450
SeniorCitizen	No vs Yes	0.818
StreamingMovies	No vs Yes	0.671
StreamingMovies	No internet service vs Yes	.
StreamingTV	No vs Yes	0.754
StreamingTV	No internet service vs Yes	.
TechSupport	No vs Yes	1.289
TechSupport	No internet service vs Yes	.

Figure 10 Odd Ratio Estimates in Output panel.

In the Output window we can also check the *Odds Ratio Estimates*. This tells us about the churn ratio of a variable by values. For example, customer with a *Contract* of *Month-to-Month* has a *5.360* times probability of churning vs a customer with a *Contract of Two year*.

Overall Accuracy

To calculate the overall accuracy, we need to construct a confusion matrix based on actual and predicted values derived from the *Classification char: Churn* plot.

Table 1 Confusion matrix for Regression model results.

		Actual Values	
Predicted Values		Positive (Churn = Yes)	Negative (Churn = No)
	Positive (Churn = Yes)	13.35227 (True Positive)	13.25758 (False Positive)
	Negative (Churn = No)	7.386364 (False Negative)	66.00379 (True Negative)

The Overall accuracy is calculated as:

$$Overall\ Accuracy: \frac{Correctly\ classified\ values}{Total\ number\ of\ values}$$

Correctly classified values: 13.35227 + 66.00379 = 79.35606

Total number of values: 100

Overall Accuracy = 0.7935606

Misclassification rate for Churn = Yes:

False Positive / True Positive + False Positive = 0.4982207716315575

Misclassification rate for Churn = No:

False Negative / True Negative + False Negative = 0.1006451628375109

2. <u>Decision Tree</u>

In the decision tree model, I have kept all setting by default. After running the model, I see *Score Rankings Overlay: Churn, Fit Statistics, Leaf Statistics, Output, Tree*, and *Treemap Plot* in the results window.

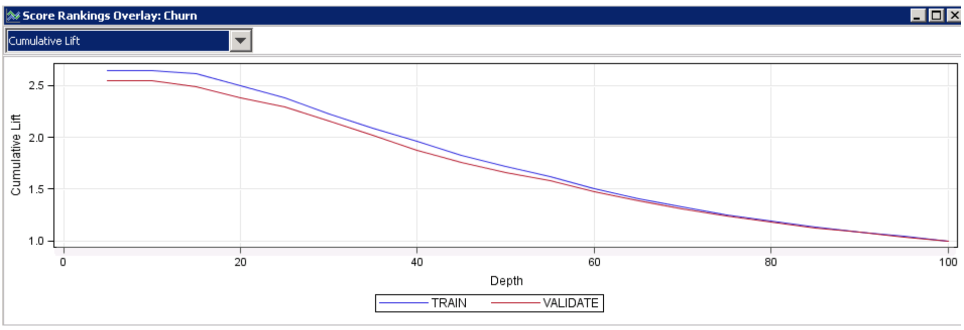


Figure 11 Score Rankings Overlay: Churn plot showing cumulative lift vs depth.

The *Score Rankings Overlay: Churn* show the *Cumulative Lift* of around *2.5* which means that this model can approximately predict *2.5* times better than a random guess for 30% of the data (till Depth = 30). If we increase the number of observations, the cumulative lift drops and will edge off to value one when we take all observations (Depth = 100) into consideration or there is no difference in the model prediction and a random guess.

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation
Churn		_NOBS_	Sum of Frequencies	4920	2112
Churn		_MISC_	Misclassification Rate	0.19878	0.208807
Churn		_MAX_	Maximum Absolute Error	0.935834	0.935834
Churn		_SSE_	Sum of Squared Errors	1369.577	621.3472
Churn		_ASE_	Average Squared Error	0.139185	0.147099
Churn		_RASE_	Root Average Squared...	0.373075	0.383535
Churn		_DIV_	Divisor for ASE	9840	4224
Churn		_DFT_	Total Degrees of Free...	4920	.

Figure 12 Fit Statics table showing statistic measures.

The *Fit Statics* window show the main statistics to determine the performance of the decision tree model. We can see that the *Misclassification Rate* for *Validation* dataset is *0.208807,* the *Average Squared Error* is *0.147099.*

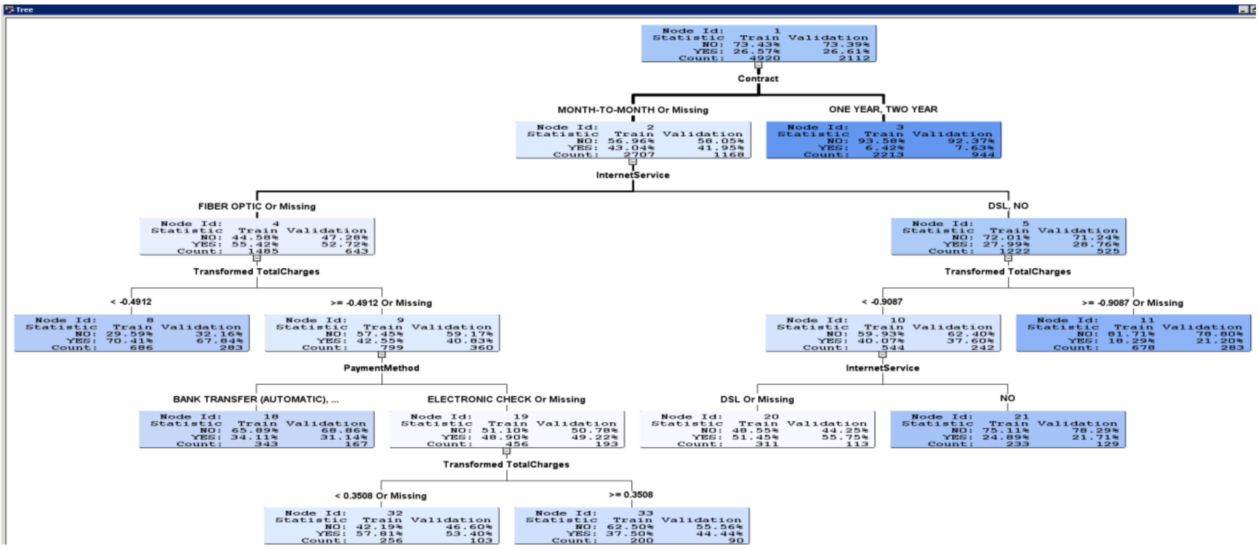


Figure 13 Decision tree in tree panel.

The *Tree* panel shows the decision tree constructed using the most important variables. As we can see Node 3 shows that customers with a *one or a two-year contract* are more likely to not churn (*Validation No: 92.37%*) as compared to customers with a *MONTH-TO-MONTH contract*. The darkness of the node tells us that the Node 3 has the highest confidence level. The *Treemap* plot shows the same information from the *Tree* plot in a concise manner.

<u>Overall Accuracy</u>

Table 2 Confusion matrix for decision tree prediction results.

<table>
<tr><td></td><td colspan="3">Actual Values</td></tr>
<tr><td rowspan="3">Predicted Values</td><td></td><td>Positive
(Churn = Yes)</td><td>Negative
(Churn = No)</td></tr>
<tr><td>Positive
(Churn = Yes)</td><td>14.67803
(True Positive)</td><td>11.93182
(False Positive)</td></tr>
<tr><td>Negative
(Churn = No)</td><td>8.948864
(False Negative)</td><td>64.44129
(True Negative)</td></tr>
</table>

The Overall accuracy is calculated as:

$$Overall\ Accuracy: \frac{Correctly\ classified\ values}{Total\ number\ of\ values}$$

Correctly classified values: 14.67803 + 64.44129 = 79.11932

Total number of values: 100

Overall Accuracy = 0.7911932

Misclassification rate for Churn = Yes:

False Positive / True Positive + False Positive = 0.4483986193082637

Misclassification rate for Churn = No:

False Negative / True Negative + False Negative = 0.1219354846973069

3. <u>Neural Network</u>

In the neural network model, I have kept all setting by default. After running the model, I see *Score Rankings Overlay: Churn, Fit Statistics, Iteration Plot*, and *Output* in the results window.

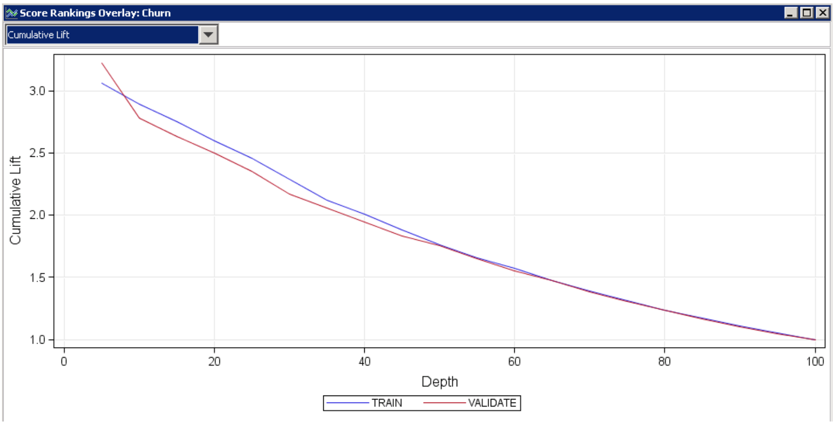


Figure 14 Score Rankings Overlay: Churn plot showing cumulative lift vs depth.

The *Score Rankings Overlay: Churn* show the *Cumulative Lift* of just above three, which means that this model can approximately predict *around three* times better than a random guess for 20% of the customers (till Depth = 20). If we increase the number of observations, the cumulative lift drops and will edge off to value one. When we take all observations (Depth = 100) into consideration or there is no difference in the model prediction and a random guess.

Fit Statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation
Churn		_DFT_	Total Degrees of Free...	4920	.
Churn		_DFE_	Degrees of Freedom f...	4847	.
Churn		_DFM_	Model Degrees of Free...	73	.
Churn		_NW_	Number of Estimated ...	73	.
Churn		_AIC_	Akaike's Information C...	4122.556	.
Churn		_SBC_	Schwarz's Bayesian Cr...	4597.134	.
Churn		_ASE_	Average Squared Error	0.131158	0.137202
Churn		_MAX_	Maximum Absolute Error	0.995591	0.995928
Churn		_DIV_	Divisor for ASE	9840	4224
Churn		_NOBS_	Sum of Frequencies	4920	2112
Churn		_RASE_	Root Average Squared...	0.362157	0.370408
Churn		_SSE_	Sum of Squared Errors	1290.592	579.5411
Churn		_SUMW_	Sum of Case Weights ...	9840	4224
Churn		_FPE_	Final Prediction Error	0.135108	.
Churn		_MSE_	Mean Squared Error	0.133133	0.137202
Churn		_RFPE_	Root Final Prediction E...	0.367571	.
Churn		_RMSE_	Root Mean Squared Er...	0.364874	0.370408
Churn		_AVERR_	Average Error Function	0.404122	0.421228
Churn		_ERR_	Error Function	3976.556	1779.265
Churn		_MISC_	Misclassification Rate	0.188618	0.196023
Churn		_WRONG_	Number of Wrong Cla...	928	414

Figure 15 Fit Statics table showing statistic measures.

The *Fit Statics* window show the main statistics to determine the performance of the decision tree model. We can see that the *Misclassification Rate* for *Validation* dataset is *0.196023,* the *Average Squared Error* is *0.137202.*

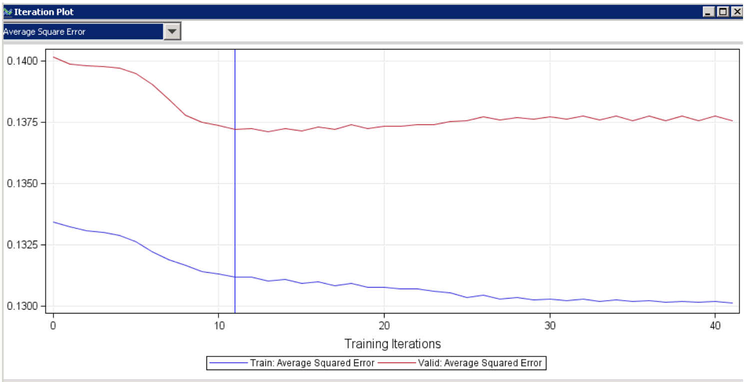


Figure 16 Iteration plot showing the ASE vs iterations.

The *Iteration Plot* shows the *Training Iterations* vs the *Average Square Error*. It shows that after eleventh iteration there is no improvement in the Average Square Error so the maximum number of iteration suitable for this model is eleven.

Overall Accuracy

Table 3 Confusion matrix for Neural Network prediction results.

		Actual Values	
Predicted Values		Positive (Churn = Yes)	Negative (Churn = No)
	Positive (Churn = Yes)	13.02083 (True Positive)	13.58902 (False Positive)
	Negative (Churn = No)	6.013258 (False Negative)	67.37689 (True Negative)

The Overall accuracy is calculated as:

$$Overall\ Accuracy: \frac{Correctly\ classified\ values}{Total\ number\ of\ values}$$

Correctly classified values: 13.02083 + 67.37689 = 80.39772

Total number of values: 100

Overall Accuracy = 0.8039772

Misclassification rate for Churn = Yes:

False Positive / True Positive + False Positive = 0.5106763097123809

Misclassification rate for Churn = No:

False Negative / True Negative + False Negative = 0.0819354935760587

3.2.4 Comparison of Models

Table 4 All three models with statistic machine learning measures.

Model	Average Squared Error	Overall Accuracy	Overall Misclassification Rate	Misclassification Rate for Churn = Yes	Misclassification Rate for Churn = No
Regression	0.141715	0.7935606	*0.206439*	0.498	0.100
Decision Tree	0.147099	0.7911932	*0.208807*	0.448	0.121
Neural Network	0.137202	0.8039772	*0.196023*	0.510	0.081

In the above table we can compare various machine learning matrices which are obtained in the previous part of the analysis. The table shows that the *Neural network* model has the lowest *Average Squared Error*, followed by *Regression* models and then *Decision Tree*. The *overall accuracy* shows the same order. The *Neural Network* has the highest accuracy, followed by *Regression*, and *Decision Tree* models. The *Overall Misclassification* is *1- Overall Accuracy*, so it shows similar results. In the *Misclassification by group (Yes/No)*, we can see *Decision Tree* has the lowest *misclassification rate for Churn = Yes* customers *(0.448)*, whereas *Neural Network* performs much better for *Churn = No* customers (with a very low misclassification rate *0.081*).

Finally, to compare all three models in SAS Enterprise Miner, I have used the *Model Comparison* node in analysis. I have set the *Selection Statistic* as *Average Square Error*. This setting will select the best model based on the lowest *Average Square Error* values. In the *Results* window for *Model Comparison* node there are four panels; *ROC Churn: Chart, Score Rankings Overlay: Churn, Output, Fit Statistics.*

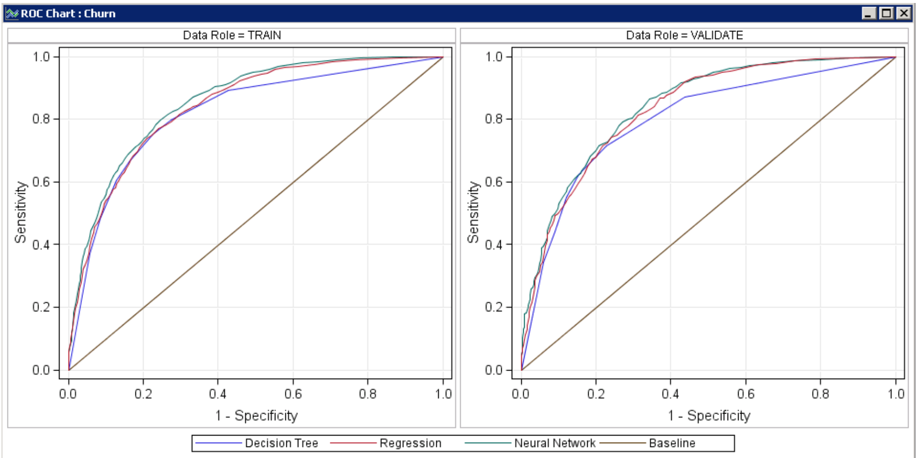


Figure 17 ROC Chart showing ROC curve for all three models.

The *ROC Chart: Churn* shows the *Specificity* vs *Sensitivity* plot. The specificity is the True Negative/ True Negative + False Positive, whereas sensitivity is True Positive/ True Positive + False Negative. So, sensitivity reflects the probability of a churn customer among customers are who are churned and are misclassified as not churned. The specificity reflects the probability of a non-churn customer among customers are who are not churned and are misclassified as churned. The ROC curve shows that the *Neural Network* (in green line) has the highest *AUC (Area under the curve)*, followed by *Regression model* (in red line), and *Decision Tree* (in blue line). This confirms our findings in the analysis performed previously.

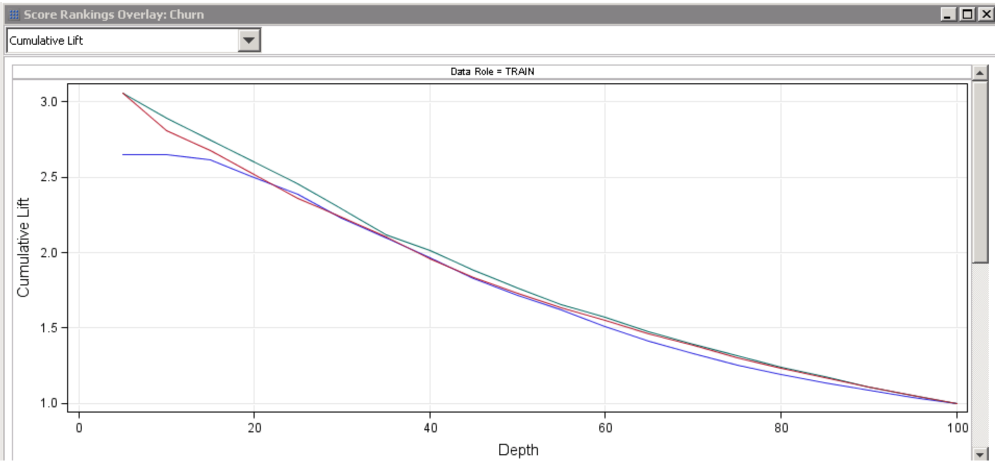


Figure 18 Score Rankings Overlay: Churn plot showing cumulative lift vs depth.

The *Score Rankings Overlay: Churn* panel shows the *Cumulative Lift* for all three models. As we can see the *Neural Network* has the highest *cumulative lift* with increasing depth. The *Regression* model has the same *cumulative lift* as *Neural Network* initially, but it drops after reaching 10% depth. We can see *Decision Tree* has the lowest and a flat *cumulative lift* initially but shows almost similar downward trend after reaching 20% depth.

Fit Statistics

Selected Model	Predecessor Node	Model Node	Model Description	Target Variable	Target Label	Selection Criterion: Valid: Average Squared Error ▲	Train: Akaike's Information Criterion	Train: Average Squared Error	Train: Average Error Function	Train: Degrees of Freedom for Error	Train: Model Degrees of Freedom	Train: T Degree Freedor
Y	Neural	Neural	Neural Net...	Churn		0.137202	4122.556	0.131158	0.404122	4847	73	
	Reg	Reg	Regression	Churn		0.141715	4173.431	0.136455	0.420267	4901	19	
	Tree	Tree	Decision Tr...	Churn		0.147099	.	0.139185	.	.	.	.

Figure 19 Fit statistics showing final selected model: Neural Network.

In the *Fit Statistics,* we can see that Neural Network is selected as the best model out of three models used in our analysis. This is based on the *Average Squared Error (ASE)* which is lowest for *Neural Network* model *(0.137202)*. The *Regression model* is the second-best model with *ASE: 0.141715*, followed by *Decision Tree Model* with *ASE: 0.147099.*

So, from the information gained we can say that *Neural Network* is the best model out of all three models used in our analysis. The selected model has more than 80 % accuracy to predict whether a customer will churn or not based on its attributes. It has a misclassification rate of around 50% for people who churn and 8% for people who do not churn. This means that the model has half a probability of misclassifying a churned customer and only 8 percent probability of misclassifying a non-churned customer. We can also confirm our results from the *Event Classification Table* in the *Output* panel in *figure 20*.

4. Campaign recommendations based on insights obtained

Customer retention is one of biggest revenue drivers for most companies are described in this article *(Plaksij, 2021)*. More than 80 percent of revenue comes from 20 percent of customers or the loyal customers. These loyal customers are our top priority in customer retention campaign. We are targeting three groups of customers: Churned, non-churned and loyal customers. Our campaign recommendations are designed around these three groups of customers.

Loyal Customer:

As we have discovered in our analysis that loyal customer have a two year contract, like fiber optic internet service, paperless billing, paying via bank transfer or credit card (automatics), opt for streaming movies and t.v, take techsupport. They have a monthly charge between 68-116 dollars, have a partner, stay more than five years, and mostly pay high for the services.

So with this information we can recommend company to invest more in fiber optic internet service, improve service in paperless billing, bank transfer/credit card (automatic), streaming movies and t.v. This will keep the loyal customers happy and satisfied with what they need. Also as they are most valueable customer and are likey to stay longer, making sure to provide them best technical support will help retaining these customers even longer. We can also offer family plans for these customer as they have partners and may like to share service with other members of the family.

<u>Non-Churned cutomers</u>

As we have discovered from our analysis, that non-churned group have a month by month and two year contract, DSL internet service, paperless billing, monthly charge between 46-110 dollars, have a partner, opt for streaming movies and t.v, don't take tech-support, stay between an year and a half, and mostly pay less for the services.

So, with this information we can recommend company to provide incentives and discount to these customer as they are likely to stay if they get more benefits. Also, company can target these customers for cold calling to offer benefits. The target should be to upgrade their internet service from DSL to fiber optics as they are more likely to turn into loyal customers if they use fiber optic services.

<u>Churned customers</u>

As we have discovered from our analysis, that churned customers have a month by month contract, fiber optics internet service, paperless billing, monthly charge between 69-105 dollars, don't have a partner, pay via electonic check, don't opt for streaming movies and t.v, don't take tech-support, stay for an year, and pay very less for the services.

So, with this information we can recommend company to talk to those who are deciding to leave, and in this way prevent customer churn. Customers can be retained if they are infomred about the competitive advantages to stay. Also we can offer long term contracts rather than a month-to-month contract. In this way customers will have enough time to use the services and see the benefits of using it. And once they see the benefits, they are more likely to commit to the product. We can also Set up milestones and offer incentives. This is a great way to make service more interactive is to set up milestones along the way. This encourages users to take the next step and continue their contract, thus it'll reduce churn rate. Also, giving incentives like discounts on upgrades or a free month of service after a certain timeframe can boost longevity *(Patel, n.d.)*.

<u>Customer recommendations based on group churn rate</u>

From our analysis we can see the groups with highest churn rates are; *SeniorCitizen (Yes), InternetService (Fiber optic), TechSupport (No), Contract (Month-to-month), and PaymentMethod (Electronic check)*. We can redesign our customer services to reduce the churn rate for senior citizens by provind more technical support. We can target to upgrade cutomers from a month-to-month contract to a long term contract. Also, we can offer customers to switch from Electronic check to credit or bank payments automatic by offering paackage discounts.

5. Conclusion

We can see from the analysis that customers who reflect attributes like month-to-month contract or with no partners are at high risk of churning. Also, Saturn Telecommunication's customer retention campaign should focus on some services more, as these services are worth investing so that more people can be satisfied and opt for long-contracts. This analysis also tells that neural network prediction model gives us better prediction results, to what we get in regression and decision tree models.

6. References

1. Patel, N. (n.d.). 4 Ways to Reduce Churn With Email Campaigns. Neilpatel.Com. Retrieved October 25, 2021, from https://neilpatel.com/blog/reduce-churn-with-email-campaigns/
2. Plaksij, Z. (2021, May 4). Customer Churn: 12 Strategies to Stop Churn Right Now! Https://Www.Superoffice.Com/. Retrieved October 25, 2021, from https://www.superoffice.com/blog/reduce-customer-churn/

7. Appendix

This figure is from *Model Comparison* node results in *Output* panel.

Event Classification Table
Model Selection based on Valid: Average Squared Error (_VASE_)

Model Node	Model Description	Data Role	Target	Target Label	False Negative	True Negative	False Positive	True Positive
Reg	Regression	TRAIN	Churn		590	3232	381	717
Reg	Regression	VALIDATE	Churn		280	1394	156	282
Tree	Decision Tree	TRAIN	Churn		516	3151	462	791
Tree	Decision Tree	VALIDATE	Churn		252	1361	189	310
Neural	Neural Network	TRAIN	Churn		606	3291	322	701
Neural	Neural Network	VALIDATE	Churn		287	1423	127	275

Figure 20 Event Classification Table in Output panel in Model Comparison node.