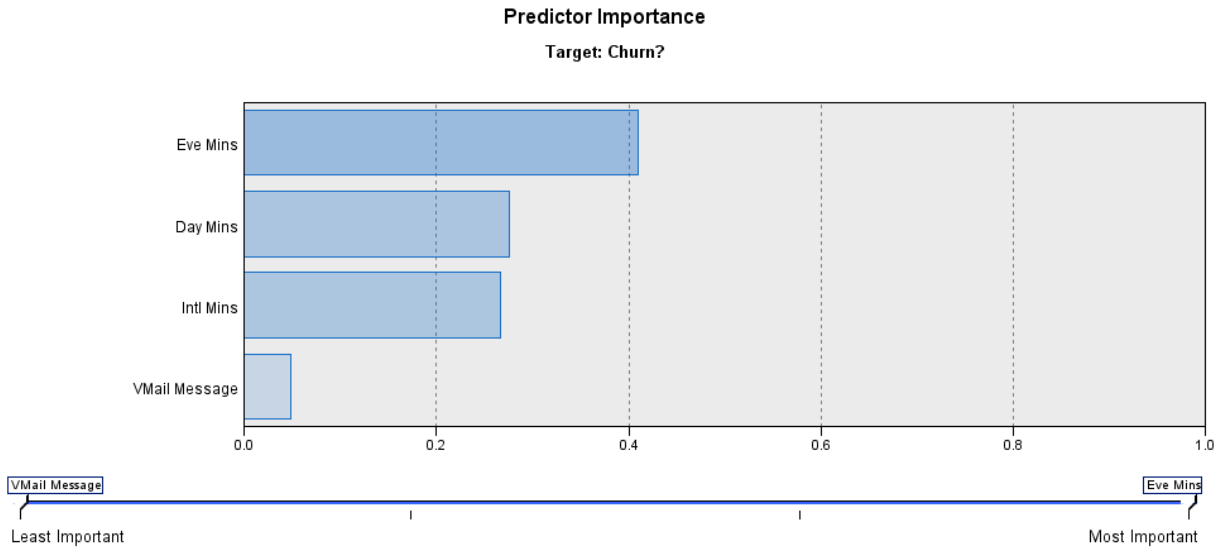
Austin C.

BIT-445-O500

Part 2: Benchmark: Logistic Regression

The way I approached the problem was to create a logistic regression model using SPSS Modeler, then to later take the data and import the Excel spreadsheet containing the Churn data, and calculate the Nagelkerke R Square results, and the correlation matrix for the predictors. In Modeler I took the data, imported it then checked the type of the data that was imported, and eliminated all inputs aside from the variables in question that could be predictors for customer churn: Vmail Message, Day Mins, Eve Mins, and IntlMins.



According to the Predictor Importance results above, it is clear that the Eve Mins variable is the most important predictor for churn out of the four variables. The least important variable is Vmail Message, with it being nearly at 0 for importance. Day Mins and Intl Mins are the two predictors that are close to being as important as Eve mins.

Variables not in the Equation

			Score	df	Sig.
Step 1	Variables	VMail Message	27.545	1	.000
		Eve Mins	29.102	1	.000
		Intl Mins	17.235	1	.000
	Overall Statistics		76.282	3	.000
Step 2	Variables	VMail Message	29.093	1	.000
		Intl Mins	17.956	1	.000
	Overall Statistics		47.699	2	.000
Step 3	Variables	Intl Mins	18.813	1	.000
	Overall Statistics		18.813	1	.000

The highest scoring variables that were not in the equation was Vmail Message and Eve Mins. The interesting thing is that Vmail Message is the least important predictor while the Eve Mins is the most important predictor according to the Predictor of Importance chart.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.253[a]	.064	.063	.341

a. Predictors: (Constant), Intl Mins, VMail Message, Day Mins, Eve Mins

R squared

This means that 6.4% of the variance in the churn can be explained. In other words this means how well all of the predictors combined can account for churn.

Classification Table[a]

Observed			Predicted		
			Churn?		Percentage Correct
			False.	True.	
Step 1	Churn?	False.	2850	0	100.0
		True.	482	1	.2
	Overall Percentage				85.5
Step 2	Churn?	False.	2850	0	100.0
		True.	477	6	1.2
	Overall Percentage				85.7
Step 3	Churn?	False.	2850	0	100.0
		True.	472	11	2.3
	Overall Percentage				85.8
Step 4	Churn?	False.	2850	0	100.0
		True.	464	19	3.9
	Overall Percentage				86.1

a. The cut value is .500

Classification table

As shown by the table, most of the churn is not accounted for by the predictors Vmail Message, Intl Mins, Day Mins, and Eve mins since there are fewer instances of there being churn caused by these predictors. 86.1% of the churn is not accounted for by the predictors in question.

Correlations

[DataSet1]

Correlations

		Day Mins	Eve Mins	VMail Message	Intl Mins
Day Mins	Pearson Correlation	1	.007	.001	-.010
	Sig. (2-tailed)		.684	.964	.558
	N	3333	3333	3333	3333
Eve Mins	Pearson Correlation	.007	1	.018	-.011
	Sig. (2-tailed)	.684		.311	.524
	N	3333	3333	3333	3333
VMail Message	Pearson Correlation	.001	.018	1	.003
	Sig. (2-tailed)	.964	.311		.869
	N	3333	3333	3333	3333
Intl Mins	Pearson Correlation	-.010	-.011	.003	1
	Sig. (2-tailed)	.558	.524	.869	
	N	3333	3333	3333	3333

Correlation Matrix

This matrix shows the correlation of each variables to one another, in regards to churn. It is noticeable that the correlation between Vmail Message and Eve mins is stronger than the others. Some correlations between variables are in the negative such as that between Eve Mins and Intl Mins or that between Day Mins and Intl Mins.