

FasterAnalytics – A Churn Case Study

DecisionQ has developed FasterAnalytics, a unique analytics package that enables researchers, analysts, and managers to use sophisticated predictive analytics from the desktop. FasterAnalytics is fast and creates high quality, predictive models from data that enable efficient review of clinical data, real-time hypothesis testing, and rapid decisions.

FasterAnalytics uses a modeling approach called Bayesian Networks to provide a mapping of the complex relationships in data, which can then be used to make high quality predictions. Users can:

- Get an instant global view of their data.
- Understand the driving factors in the data.
- Test hypotheses in real time in our model Explorer.
- Produce reports that can be exported to other applications.
- Make determinations that can help prioritize the use of scarce research resources.

Market Overview

Product, Marketing and Sales Managers consistently desire better tools to enhance their sales efforts. FasterAnalytics can help organizations turn consumer and financial data into valuable information, allowing them to make decisions with advanced knowledge of high-probability consumer and market reactions. To assist in this effort, DecisionQ has developed FasterAnalytics for Consumer Products & Financial Services, a tool for modeling consumer and financial data. DecisionQ's FasterAnalytics software can combine data and business experience to create powerful predictive models, models that can be used to understand future consumer behavior.

Value to the Customer

FasterAnalytics enables both experts and non-experts in statistics to discover and leverage knowledge from large quantities of data quickly. Examples include:

- Automatically mapping data where targets are unknown to reveal correlations.
- Discovering new relationships between variables and identifying new opportunities to improve sales or reduce cost.
- Identifying potential concerns early.
- Discovering populations that may have substantially different responses from the population at large.
- Predicting the behavior of any factor or combination of factors in the model.

FasterAnalytics is designed for real-time environments. Bayesian models are highly effective at identifying emerging trends that can be used to either to identify potential adverse events or improve quality of outcomes.

Product and Technology

DecisionQ Corporation has produced a range of modules that include data analysis, modeling, visualization, reporting, and decision optimization. FasterAnalytics modules include:

- *Discretizer*. Automatically configures the data for modeling.
- *Modeler*. Quickly creates a visual model of the data.
- *Explorer*. Allows real-time generation and testing of hypotheses.

- *Reporter.* Extracts insights and key points for inclusion in reports and presentations.

Using the System: A Predictive Example

The following is an example application of our software to churn data set. We have used a set comprised of 5000 telephone customers with 21 attributes or markers. FasterAnalytics built the model in this example, from start to finish, in less than 5 minutes.

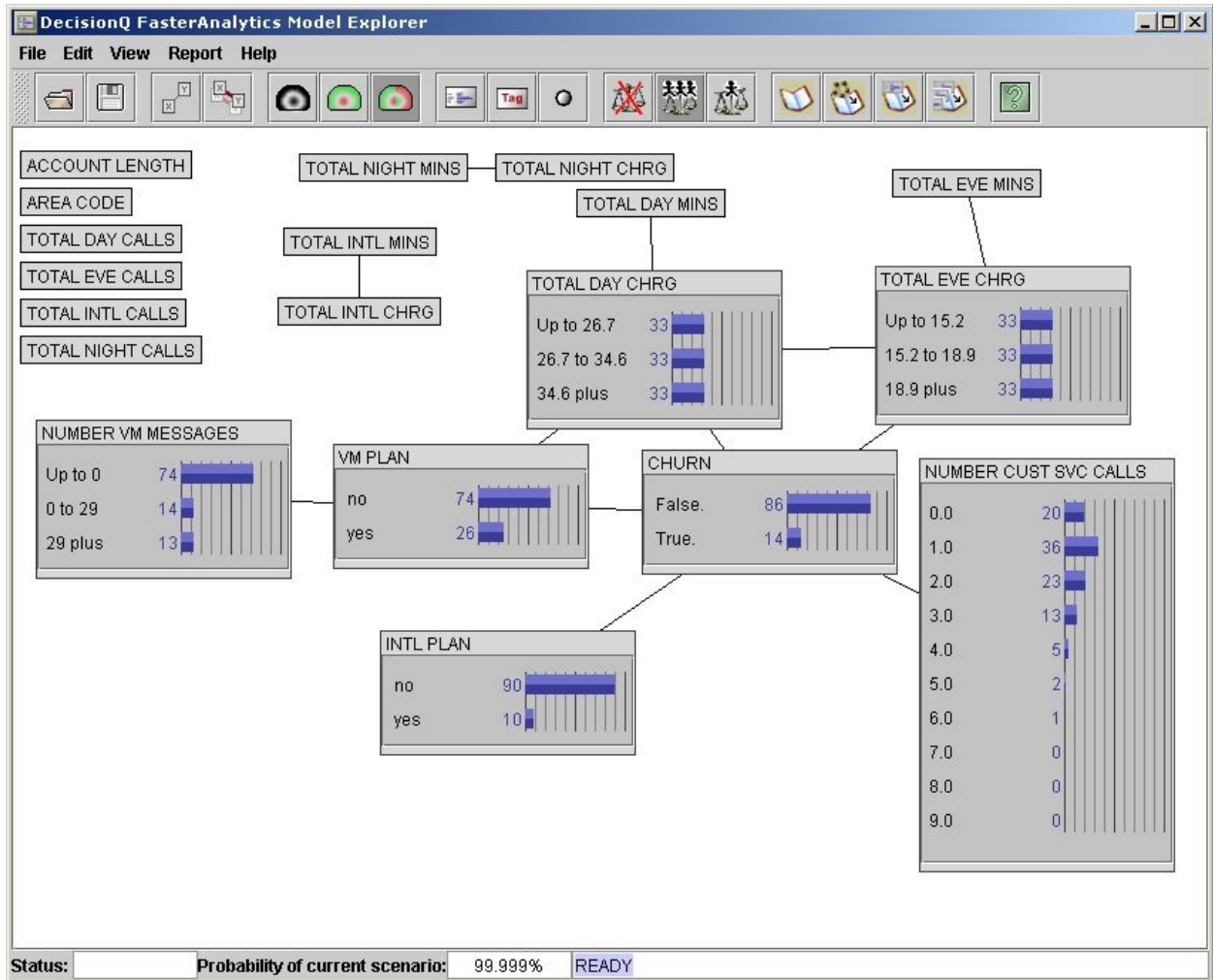
To build predictive models, our learning engine requires the data to be in a flat tabular format. The data can be numerical or variable character strings. Our software also handles missing values automatically and will either impute a value or treat missing values as a special category, at the user’s discretion.

Figure 1: This example uses a data set held in an Excel spreadsheet as shown below

	A	B	C	D	E	F	G	H	I	J
1	STATE	ACCOUNT	AREA COI	PHONE N	INTL PLAN	VM PLAN	NUMBER	TOTAL DA	TOTAL DA	TOTAL DA
2	KS	128	615	382-4657	no	yes	25	265.1	110	45.07
3	OH	107	615	371-7191	no	yes	26	161.6	123	27.47
4	NJ	137	615	358-1921	no	no	0	243.4	114	41.38
5	OH	84	608	375-9999	yes	no	0	299.4	71	50.9
6	OK	75	615	330-6626	yes	no	0	166.7	113	28.34
7	AL	118	510	391-8027	yes	no	0	223.4	98	37.98
8	MA	121	510	355-9993	no	yes	24	218.2	88	37.09
9	MO	147	615	329-9001	yes	no	0	157	79	26.69
10	LA	117	608	335-4719	no	no	0	184.5	97	31.37
11	WV	141	615	330-8173	yes	yes	37	258.6	84	43.96
12	IN	65	615	329-6603	no	no	0	129.1	137	21.95
13	RI	74	615	344-9403	no	no	0	187.7	127	31.91
14	IA	168	608	363-1107	no	no	0	128.8	96	21.9
15	MT	95	510	394-8006	no	no	0	156.6	88	26.62
16	IA	62	615	366-9238	no	no	0	120.7	70	20.52
17	NY	161	615	351-7269	no	no	0	332.9	67	56.59

Having selected the data, a fully automated process will continue until a full model is presented, or the user can stop each part of the process to manually change parameters. The software begins by categorizing the data and 'binning' in accordance with the default settings. The data is then passed seamlessly to the Modeler for automated model development. Once the software has mapped the complex correlations in the data a model is presented in the Explorer.

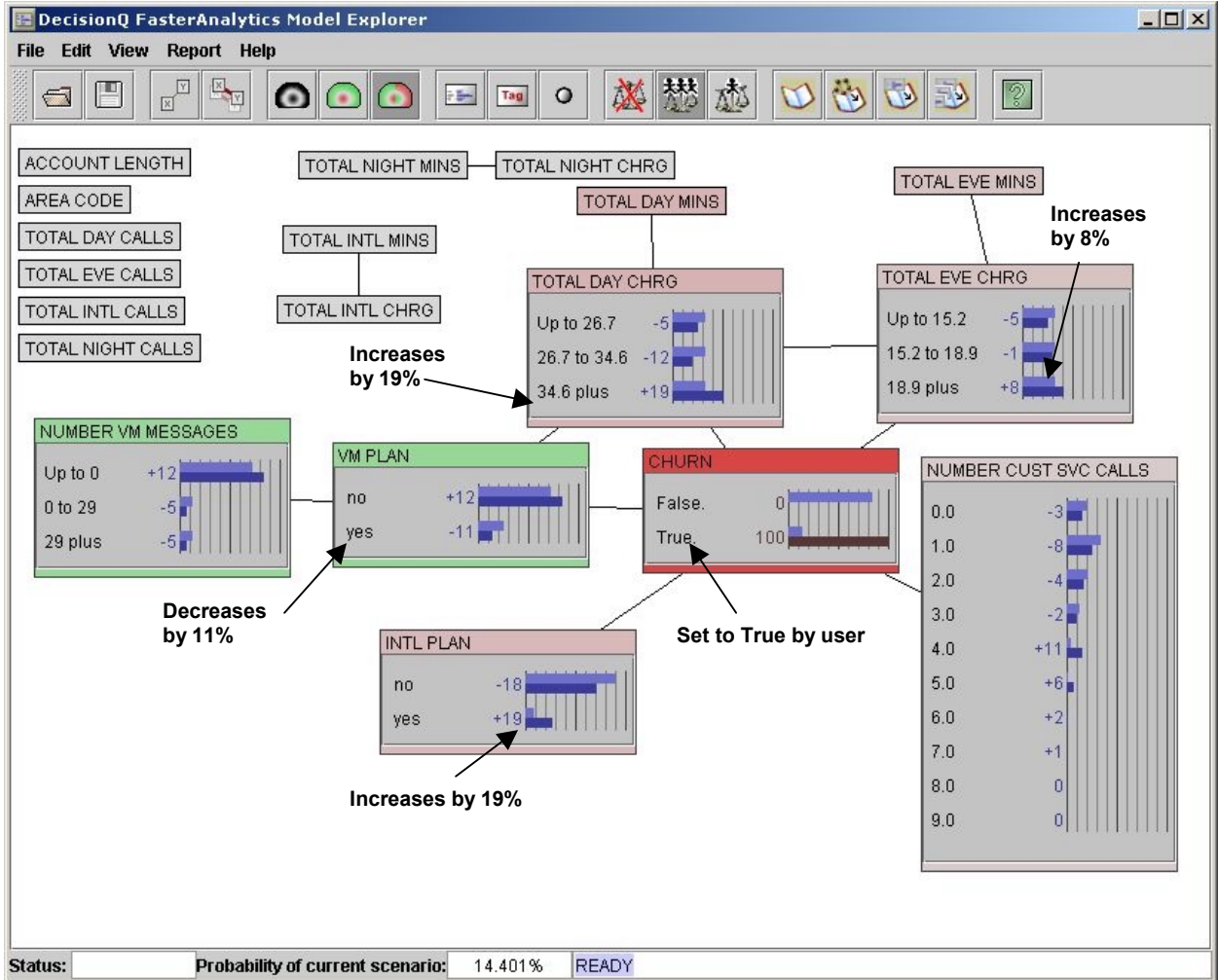
Figure 2: Base case model of the data presented in Explorer



The display illustrates conditional dependence between variables and the pathways existing in the predictive model. Notice that the network has multiple branches, and that the data is interrelated in a “web”, one of the strengths of multivariate Bayesian networks.

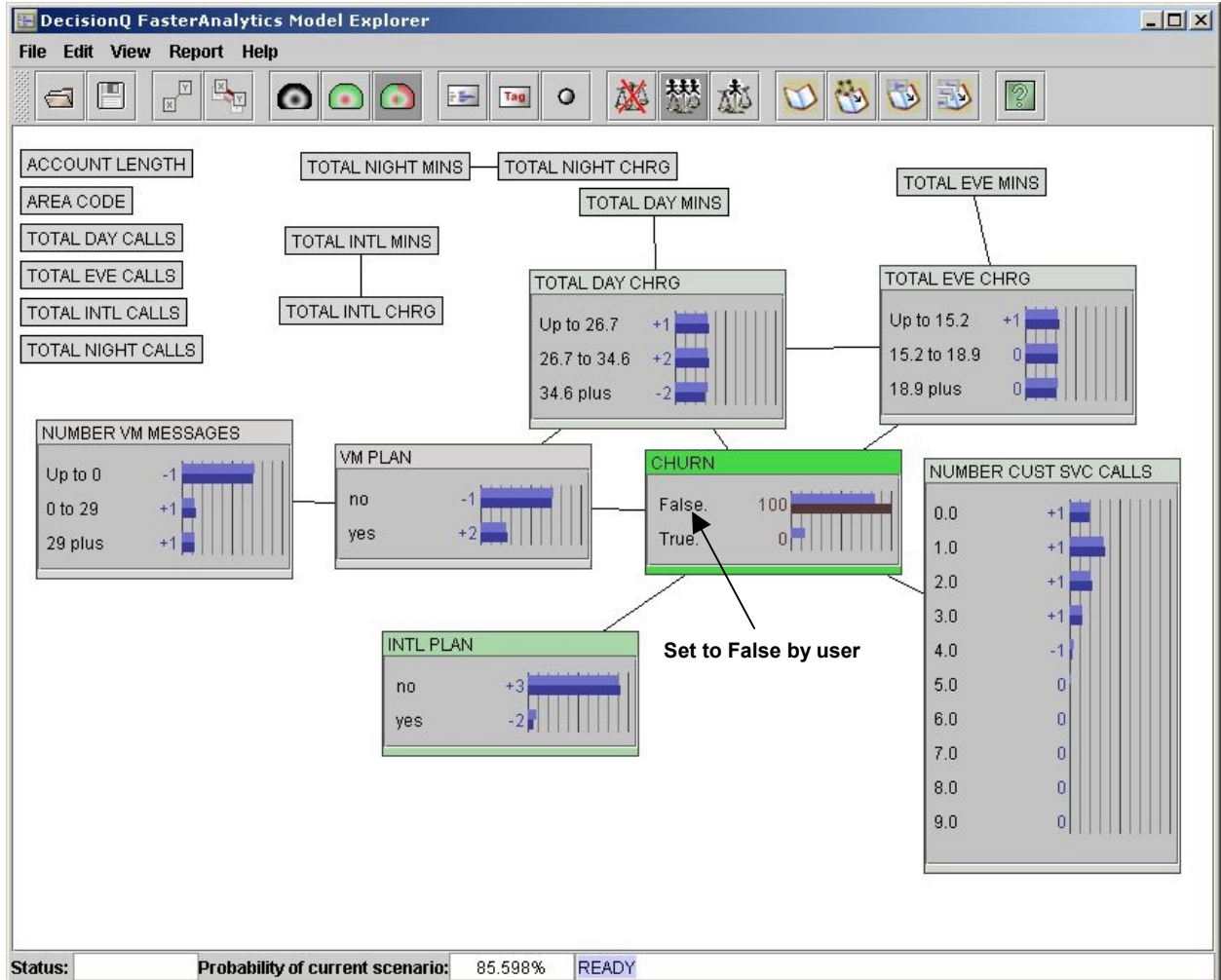
In the example below, we examine variables that correlate with churn. We begin by selecting our target variable, CHURN, and setting it to "True." We can see that the coloring and distribution of the surrounding nodes has changed to indicate the effect on other variables associated with the current case. Note that the number of customers with a voice mail plan (VM PLAN) has decreased, while the number of customers with an international dialing plan (INTL PLAN) and high levels of day (TOTAL DAY CHRG) and evening charges (TOTAL EVE CHRG) has increased.

Figure 3: CHURN set to True



Next, we set CHURN to “False” and examine the effect on the other variables. Note that our variables (VM PLAN, INTL PLAN, TOTAL DAY CHRNG, and TOTAL EVE CHRNG) retain approximately the same value.

Figure 4: CHURN set to False

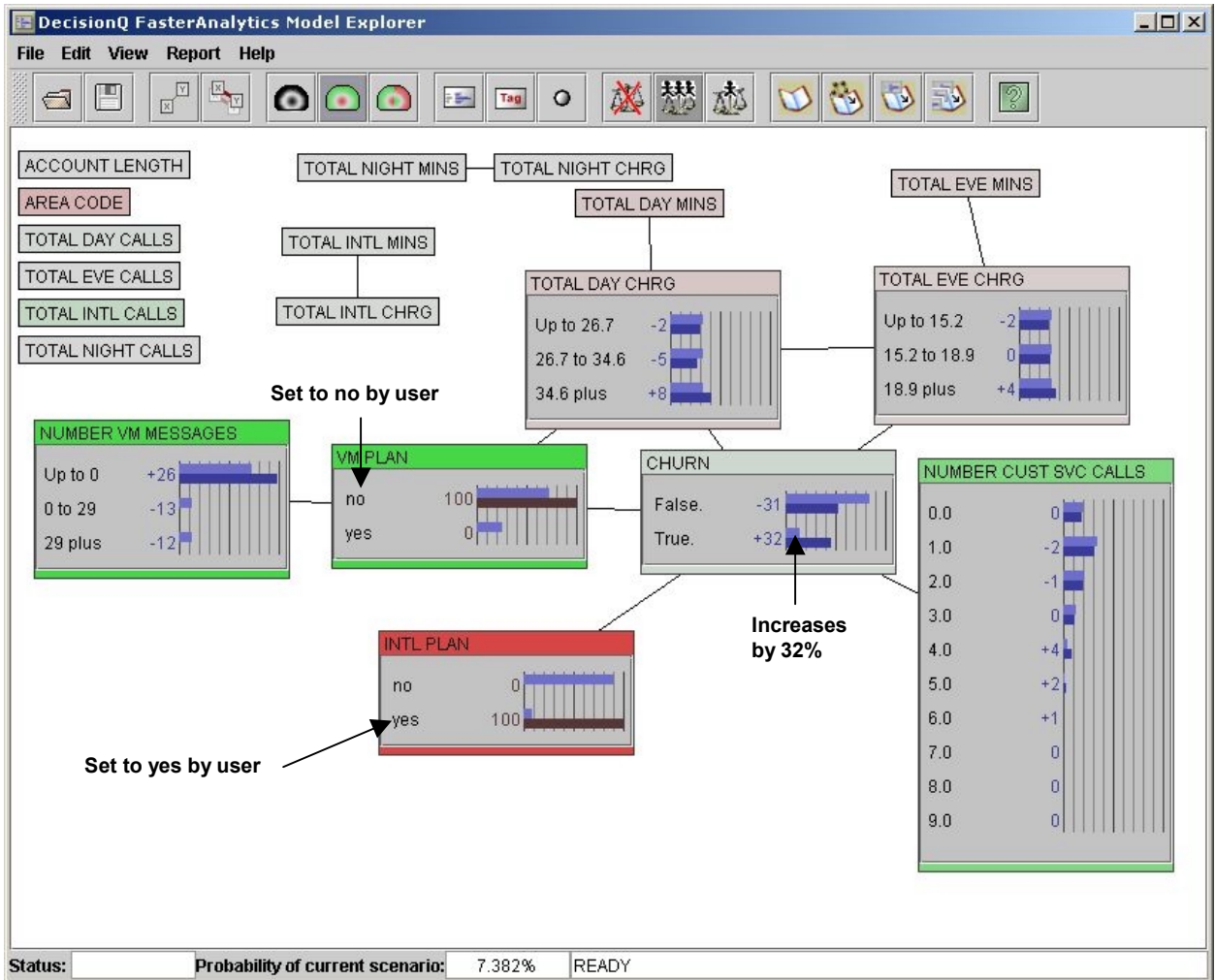


Compare the two models in Figure 3 and 4 above with the base level in Figure 2. While CHURN shares conditional dependence with the same nodes, the behavior of those markers changes based upon the presence of churn. The coloring in the graphical model shows the change in population profile quickly and effectively.

It is also possible to select two or more variables simultaneously. For example, the extent to which VM PLAN and INTL PLAN affect CHURN can be studied together. If we wish to test hypotheses, we can modify any node and see how our hypothesis affects the model. Notice how information flows through the network.

Suppose we are interested in examining how VM PLAN and INTL PLAN affect CHURN as predictive variables (Figure 5). We first select these nodes and click “Graph” to display the states within these nodes. This can be done for as many variables as we choose. We then set VM PLAN to “no” and INTL PLAN to “yes.” Here we see very clearly the predicted result for CHURN level. If a customer does not have a voice mail plan but has an international plan, the likelihood of churn increases by 32%.

Figure 5: VM PLAN combined with INTL PLAN and correlation with CHURN



The Reporter module can be used to create a report that will show the conditional probabilities (or predicted likelihood) of any target variables, given the expression of any independent variable(s). Any part of the model, visualization can be pasted into Reporter and then transferred into other applications. Figure 6 shows a sample report. Figure 7 shows the sample report pasted into a Microsoft Excel worksheet.

Figure 6: A sample report listing the relationship between the variable CHURN with the variables of VM PLAN and INTL PLAN

The screenshot shows a software window titled "DecisionQ FasterAnalytics Report". The window has a menu bar with "File" and "Edit", and a toolbar with icons for file operations and refresh. The main content area displays a report titled "DecisionQ FasterAnalytics Report" containing a table with the following data:

Probability of case	Drivers		Target	
	INTL PLAN	VM PLAN	CHURN	
			False.	True.
66.117%	no	no	86.7	13.3
7.382%	yes	no	53.1	46.9
24.338%	no	yes	93.7	6.3
2.16%	yes	yes	72.2	27.8

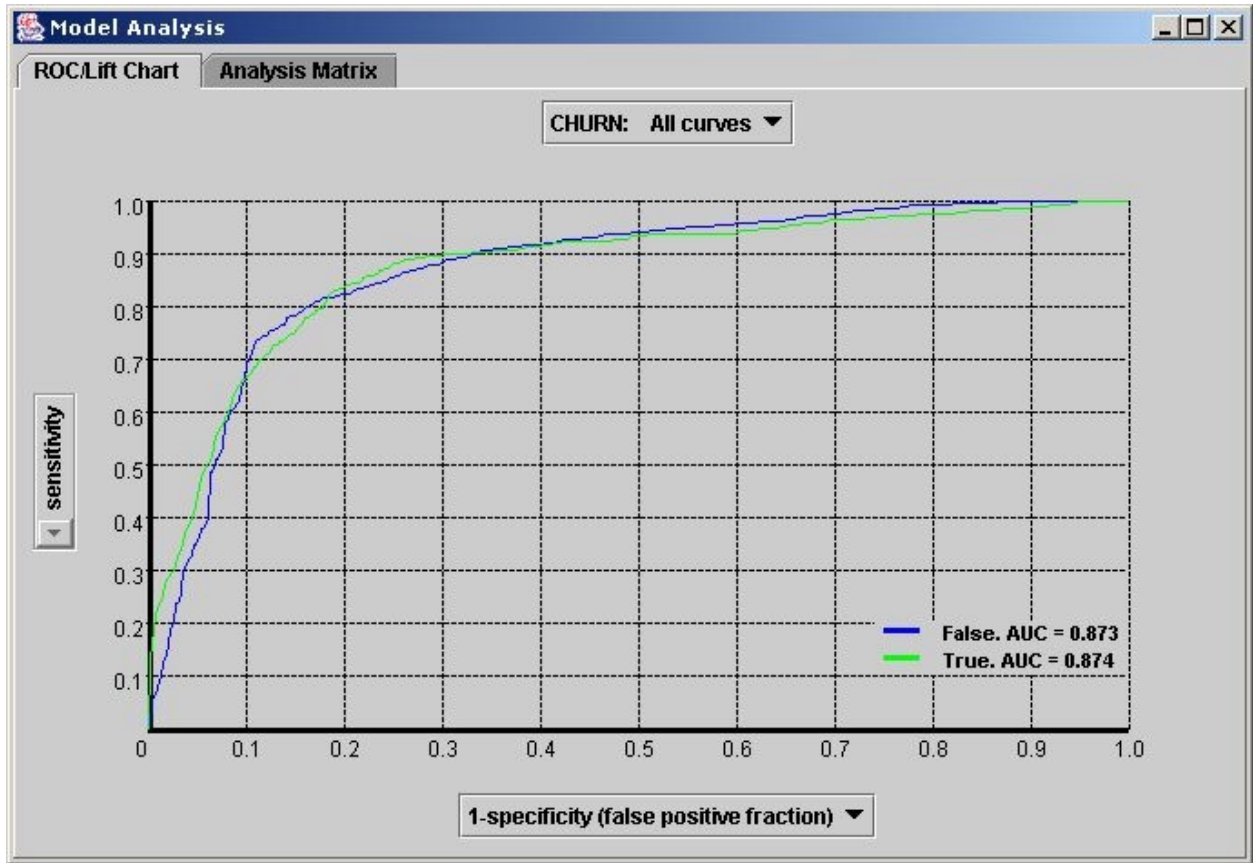
Figure 7: A sample report pasted into a Microsoft Excel worksheet

The screenshot shows a Microsoft Excel window titled 'Microsoft Excel - Book1'. The menu bar includes File, Edit, View, Insert, Format, Tools, Data, Window, and Help. The toolbar shows various icons for file operations and formatting. The active cell is F25. The worksheet contains a report with the following data:

	A	B	C	D	E	F	G
1	DecisionQ FasterAnalytics Report						
2							
3							
4	Probability	Drivers		Target			
5	of case	INTL PLAN	VM PLAN	CHURN			
6				False.	True.		
7	66.12%	no	no	86.7	13.3		
8	7.38%	yes	no	53.1	46.9		
9	24.34%	no	yes	93.7	6.3		
10	2.16%	yes	yes	72.2	27.8		
11							
12							
13							

ROC curves are a new feature added to FasterAnalytics. Accessed through the Modeler, the ROC curves estimate the predictive accuracy of the model by testing the algorithm created by the model against the data set. Figure 8, below, shows the ROC curves and AUC (area-under-curve) values for both instances of churn (True and False). These numbers indicate that the model/algorithm created by FasterAnalytics is of very high quality.

Figure 8: ROC Curves



DecisionQ sells predictive modeling software and complementary professional services. Alternatively, components from FasterAnalytics can be integrated into third party applications as part of broad data management and analysis platform. If you have any further questions or would like to schedule a more detailed demonstration in person or over the web, please contact us.

DecisionQ Corporation
3726 Connecticut Ave. NW, Suite 519
Washington, D.C. 20008
www.decisionq.com
Phone: 415-254-7996
Fax : 415-276-6356
Email: info@decisionq.com