

INTRODUCTION TO SAS® ENTERPRISE MINER™

**MELODIE RUSH
CUSTOMER LOYALTY SYSTEMS ENGINEER**



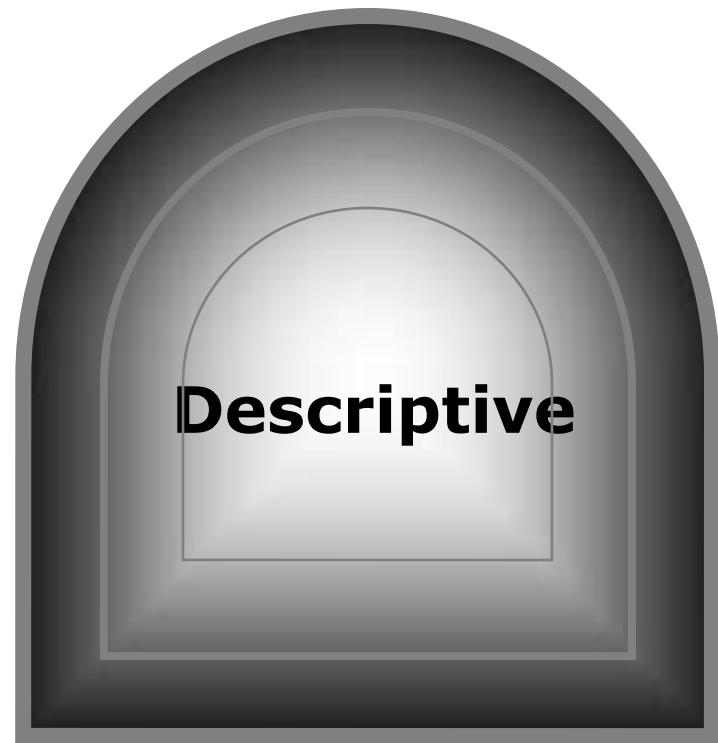
TODAY'S AGENDA

- What is Data Mining & SAS Enterprise Miner
 - Benefits of using SAS Enterprise Miner
- SAS Enterprise Miner Demonstration
 - Enterprise Miner GUI Overview
 - Creating a Data Mining Flow
 - Modeling & Model Comparison
 - Scoring in Enterprise Miner
 - Rapid Predictive Modeler (RPM)
- Q&A

WHAT IS DATA MINING?

Turning increasing amounts of
raw data into useful information

TWO PASSAGES INTO THE DATA MINE



Clustering (Segmentation)

grouping together similar people, things, events

- Transactions that are likely to be fraudulent, Customers that are likely to have similar behaviors.

Associations

affinity, or how frequently things occur together, and sometimes in what order

- Customers who purchase product A also purchase product B

Classification models

predict class membership

- 0 or 1: 1 if person responded; 0 otherwise
- Low, Medium, High: a customer's likeliness to respond

Regression models

predict a number

- \$217.56 – Total profit, expense, cost for a customer
- 37 – The number of months before a customer churns

SAS ENTERPRISE MINER

DEVELOPED TO...

- **Enhance Productivity for a Wider Target Audience**
 - Focus on analytic skills rather than coding – quicker time to analytical effectiveness
 - Team working and sharing of best practice, with self documenting and audit trail
- **Provide Additional Data Mining Algorithms and Techniques**
 - Which are not available outside EM
 - Less reliance on coding skills - encourages more exploration/use of different techniques
- **Generate Complete Scoring Formula**
 - Capture the scoring code for all stages of model development
 - Reduce potential errors that can arise from manual conversions
 - Deploy models quickly and ‘automatically’
 - Options: SAS, C, Java, PMML, “In-Database”
- **Provide Integrated Model Assessment**
 - Enables results from numerous models and modeling techniques to be compared.

EM ANALYTICAL FUNCTIONALITY INCLUDES:

- Neural Networks
- Kohonen Networks
- Self-Organising Maps
- K-Means Clustering
- Principal Component Analysis
- Rules Builder
- Interactive Grouping
- Decision Trees (CHAID, CART, C4.5, Hybrid)
- Bagging, Boosting, Ensemble
- Regression (Linear & Logistic, Stepwise, forward & backward selection)
- Variable Clustering
- Imputation
- Support Vector Machines
- Associations and Sequence Analysis
- Rule Induction
- Two Stage Modelling
- Memory Based Reasoning
- Time Series Analysis
- Text Mining
- Sampling (Random, Stratified, Weighted, Cluster ,Systematic)
- Filtering (outliers and rules)
- Variable Selection / Grouping
- Data partition, merge, append
- Assessment & Model Comparison (Lift charts, ROC curves, profit/cost matrix, Confusion matrix)
- Credit scoring functionality
- Automatic score code generation for the whole process

UNIQUE EM ANALYTICAL FUNCTIONALITY

- Neural Networks
- Kohonen Networks
- Self-Organising Maps
- (Automatic) K-Means Clustering
- Principal Component Analysis
- Rules Builder
- Interactive Grouping
- Decision Trees (CHAID, CART, C4.5, Hybrid)
- Bagging, Boosting, Ensemble
- Regression (Linear & Logistic, Stepwise, forward & backward selection)
- Variable Clustering
- Imputation
- Support Vector Machines
- Associations and Sequence Analysis
- Rule Induction
- Two Stage Modelling
- Memory Based Reasoning
- Time Series Analysis
- Text Mining
- (Out-of-the-box) Sampling (Random, Stratified, Weighted, Cluster, Systematic)
- Filtering (outliers and rules)
- Variable Selection / Grouping
- (Out-of-the-box) Data partition, merge, append
- (Automatic) Assessment & Model Comparison (Lift charts, ROC curves, profit/cost matrix, Confusion matrix)
- Credit scoring functionality
- Automatic score code generation for the whole process

(I.E. NOT IN SAS/STAT)

SAS® ENTERPRISE MINER MODEL DEVELOPMENT PROCESS

S_{ample}

- Input Data
- Sample
- File Import
- Data Partition
- Merge
- Append
- Time Series
- Filter

E_{xplore}

- Association
- Cluster
- MDDB
- Graph Explore
- Market Basket
- MultiPlot
- Path Analysis
- SOM/Kohonen
- StatExplore
- Variable Clustering
- Variable Selection

M_{odify}

- Transform Variables
- Replacement
- Impute
- Interactive Binning
- Principal Components
- Drop
- Rules Builder

M_{odel}

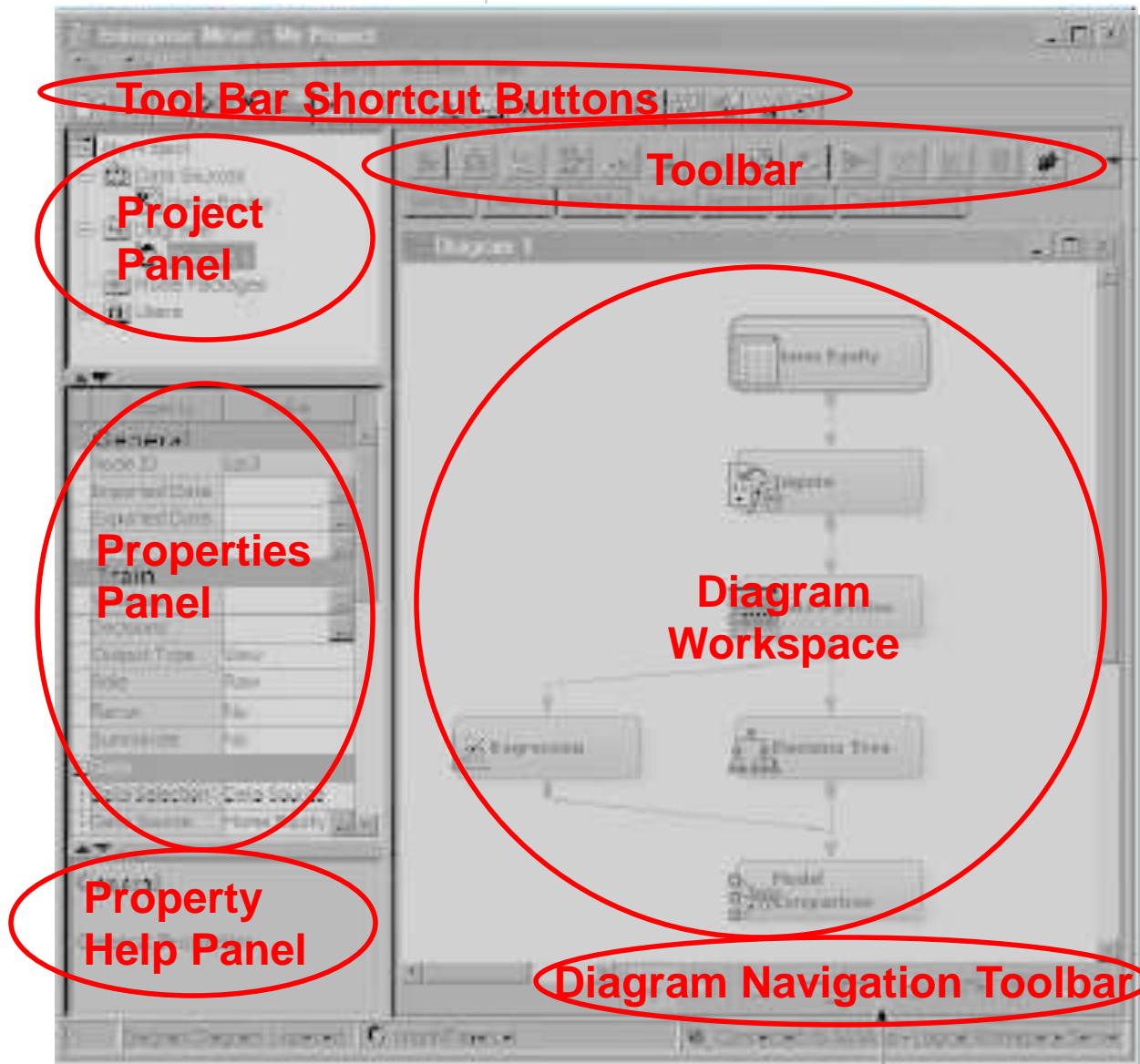
- Decision Tree
- Gradient Boosting
- AutoNeural
- Neural Network
- DMNeural
- Regression
- Dmine Regression
- MBR
- Partial Least Squares
- LARS
- Model Import
- Rule Induction (2)
- Ensemble
- TwoStage

A_{ssess}

- Model Comparison
- Score
- Segment Profile
- Decisions
- Cutoff

Utility

- Metadata
- SAS Code
- Reporter
- Start Groups
- End Groups
- Score Code Export



SAS Enterprise Miner GUI

SAS® ENTERPRISE MINER™ DEMO



SAS® RAPID PREDICTIVE MODELER

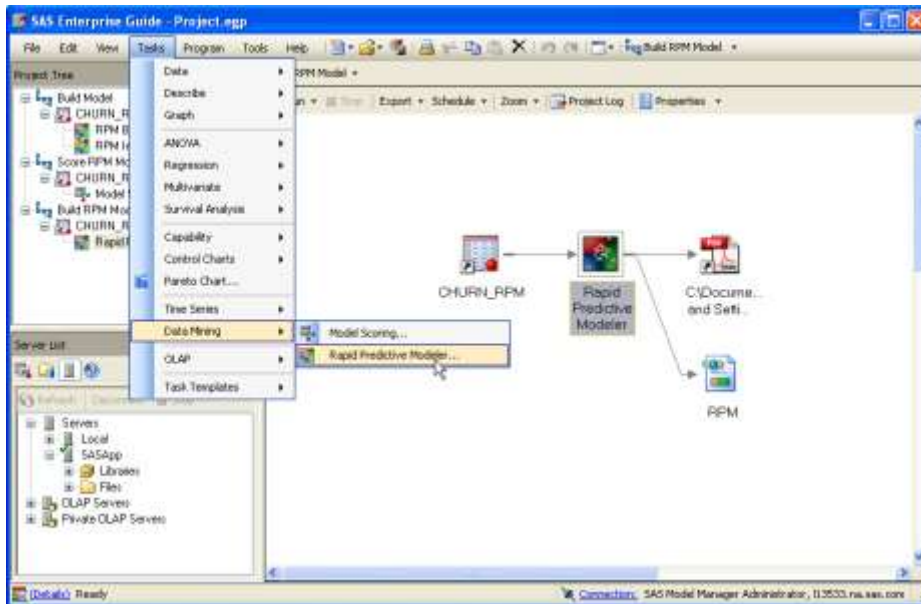


- Need to generate numerous models to solve a variety of business problems in a credible manner
- Models need to be developed in a quick time-frame using a self-service approach
- Does not want to always rely on analytic professionals (e.g. statistician or modeler or data miner)

- Solving more complex issues on hand to gain incremental value
- Further customize or refine models for better results

SAS RAPID PREDICTIVE MODELER PACKAGING

- Included in the SAS Enterprise Miner bundle at no additional charge
- Delivered as a customized task in SAS Enterprise Guide and the SAS Add-in for MS Office (MS Excel only)

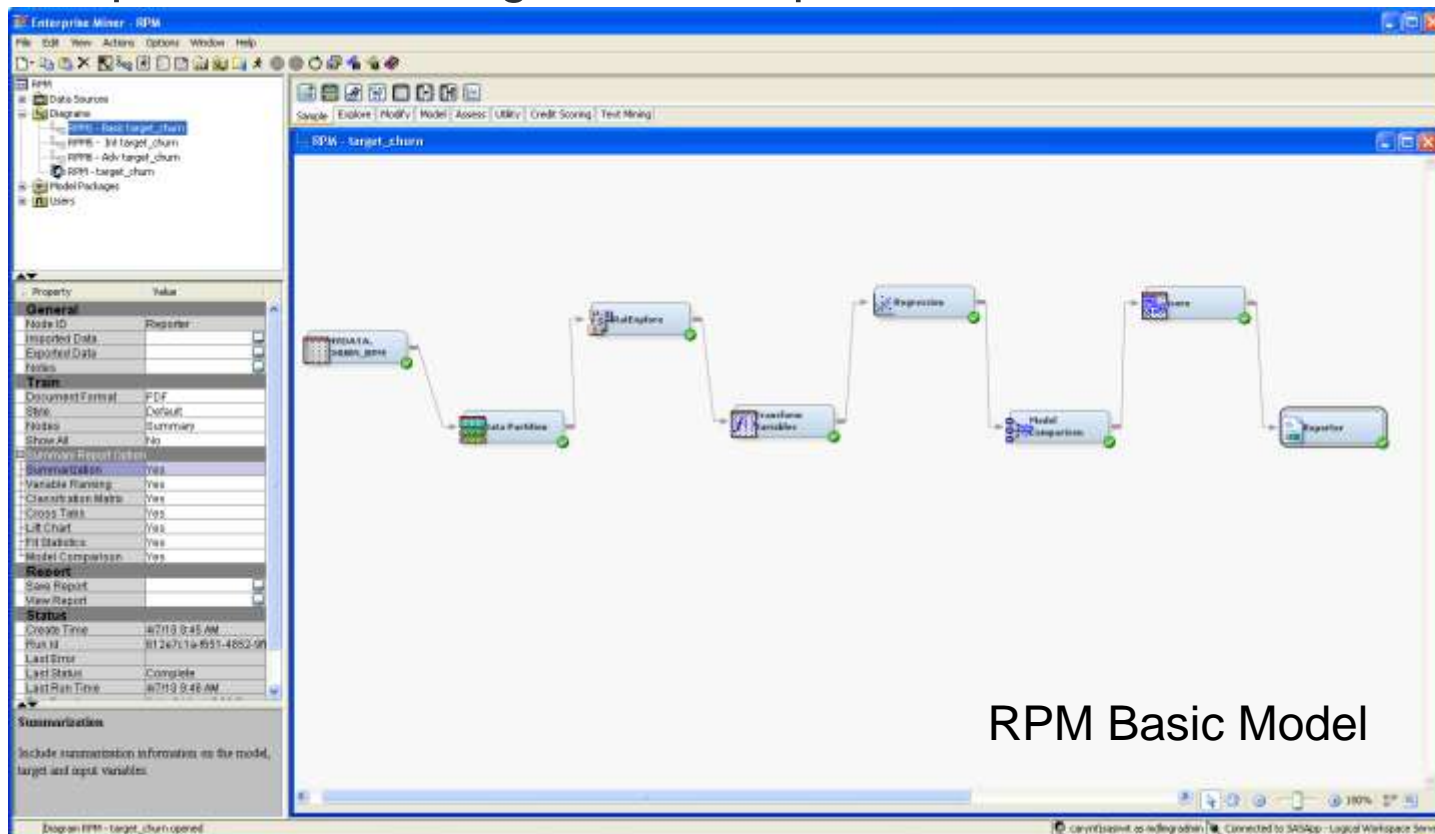


The screenshot shows the Microsoft Excel interface with the SAS Add-in for MS Office. The 'SAS' menu is open, and the 'Rapid Predictive Modeler...' option is selected. The data table is displayed, showing columns for 'churn', 'current bill amount', 'Avg. Calls', and 'Avg. Calls Weekdays'. The data is organized into rows, with the first row being the header and subsequent rows containing numerical values.

	churn	current bill amount	Avg. Calls	Avg. Calls Weekdays
1	1	14407	0	0
2	1	14030	5978.333333	2136
3	1	10594	12712	204.6666667
4	1	13807	15490.33333	5075
5	1	3805	4776	3470.333333
6	1	2847	54.66666667	5147.666667
7	1	10238	21866	2015
8	1	8082	2667	0
9	1	3347	4982.333333	6506.666667
10	1	14788	11543	326.6666667
11	1	47991	4616	4344
12	1	22061	12681.66667	34888
13	1	1101318	4034	73
14	1	1101319		
15	1	1101323		
16	1	1101503		
17	1	1101591		
18	1	1101720		
19	1			
20	1			
21	1			

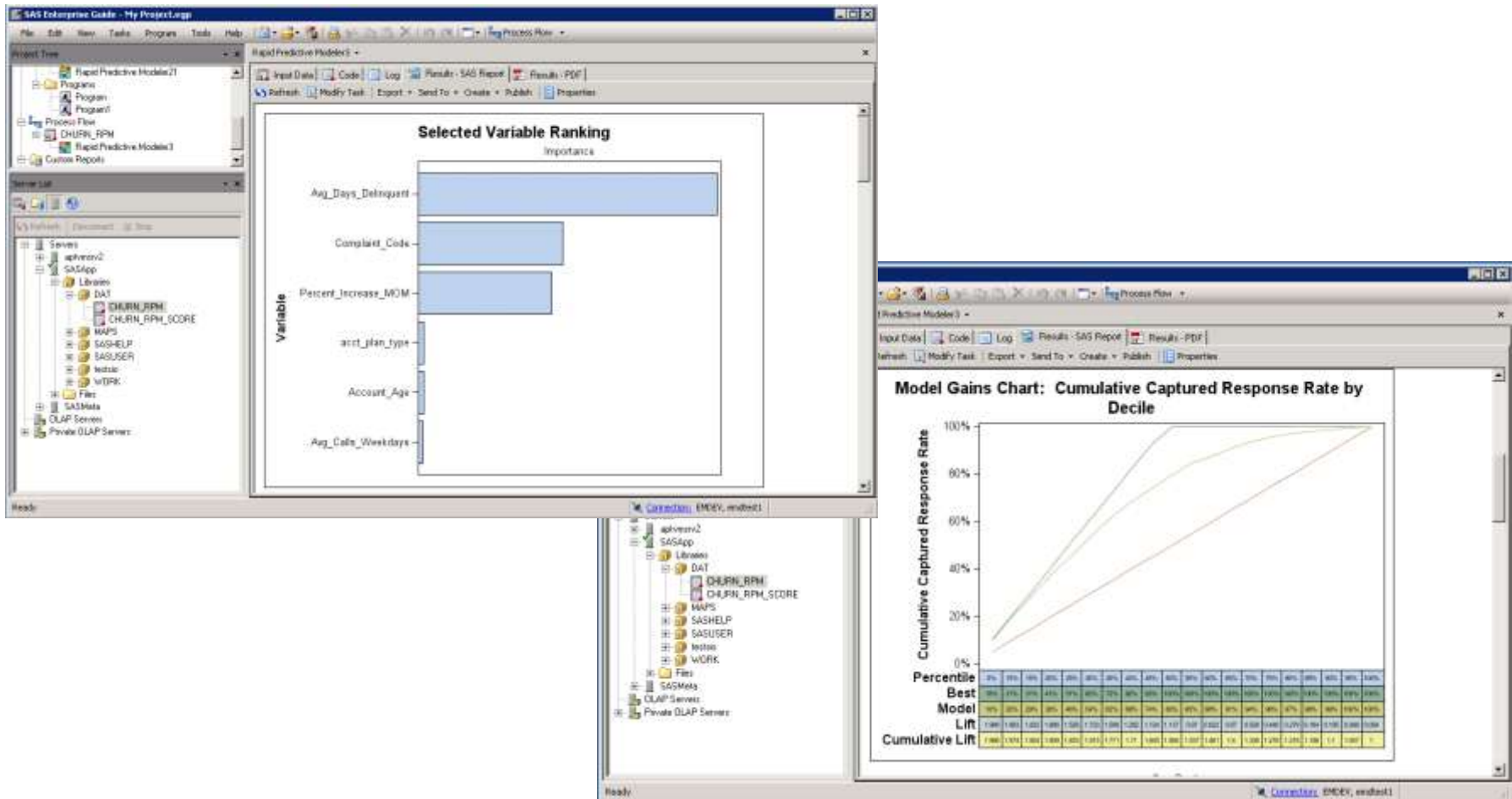
SAS RPM KEY CAPABILITIES

- Chose from prebuilt Enterprise Miner models that use a broad range of classical and modern modeling techniques.
- Analytic experts can further customize and improve SAS RPM developed models using SAS Enterprise Miner.



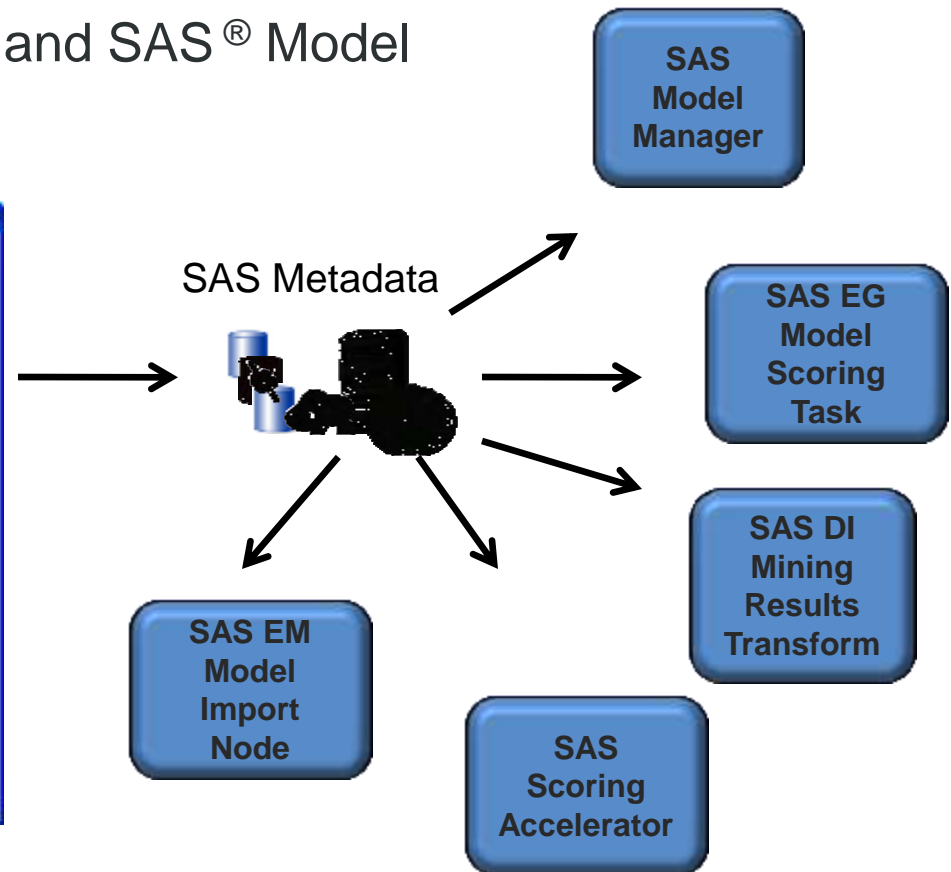
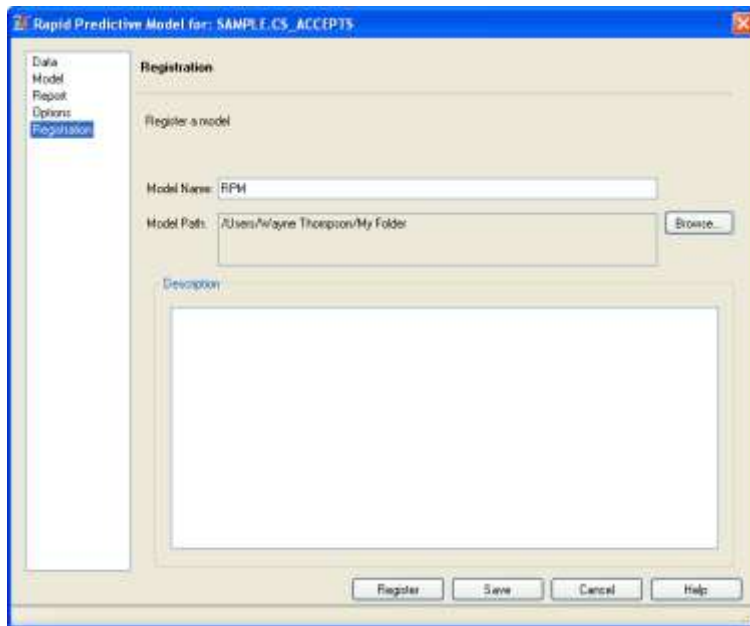
SAS RPM KEY CAPABILITIES

- Analytic results are presented as a simple to understand reports including scorecard, lift charts, and listing of key variables.



SAS RPM KEY CAPABILITIES

- Models can be registered in SAS metadata for direct use in other products such as SAS[®] Enterprise Guide, SAS[®] Data Integration Studio, and SAS[®] Model Manager.



DEMONSTRATE RAPID PREDICTIVE MODELER TASK



ENTERPRISE MINER RESOURCES

- [SAS Rapid Predictive Modeler External Website](#)
 - Product brief, Press release, Brief product demo, etc.
- E-Learning Class for Rapid Predictive Modeler (RPM)
 - [Rapid Predictive Modeling for Business Analysts](#)
- [SAS Enterprise Miner External Web Site](#)
 - [SAS Enterprise Miner Technical Support Web Site](#)
 - [SAS Enterprise Miner Technical Forum](#) (Join Today!)
 - [SAS Enterprise Miner Training](#)
 - [Getting Started with SAS Enterprise Miner](#) (scroll down to your version of EM)

FURTHER READING

- [Identifying and Overcoming Common Data Mining Mistakes](#) by Doug Wielenga, SAS Institute Inc., Cary, NC
- [Best Practices for Managing Predictive Models in a Production Environment](#) by Robert Chu, David Duling, Wayne Thompson , SAS Institute Cary, NC
- [From Soup to Nuts: Practices in Data Management for Analytical Performance](#) by David Duling, Howard Plemmons, Nancy Rausch, SAS Institute Cary, NC
- (All available on support.sas.com)

THANK YOU FOR USING SAS!



**THE
POWER
TO KNOW.**