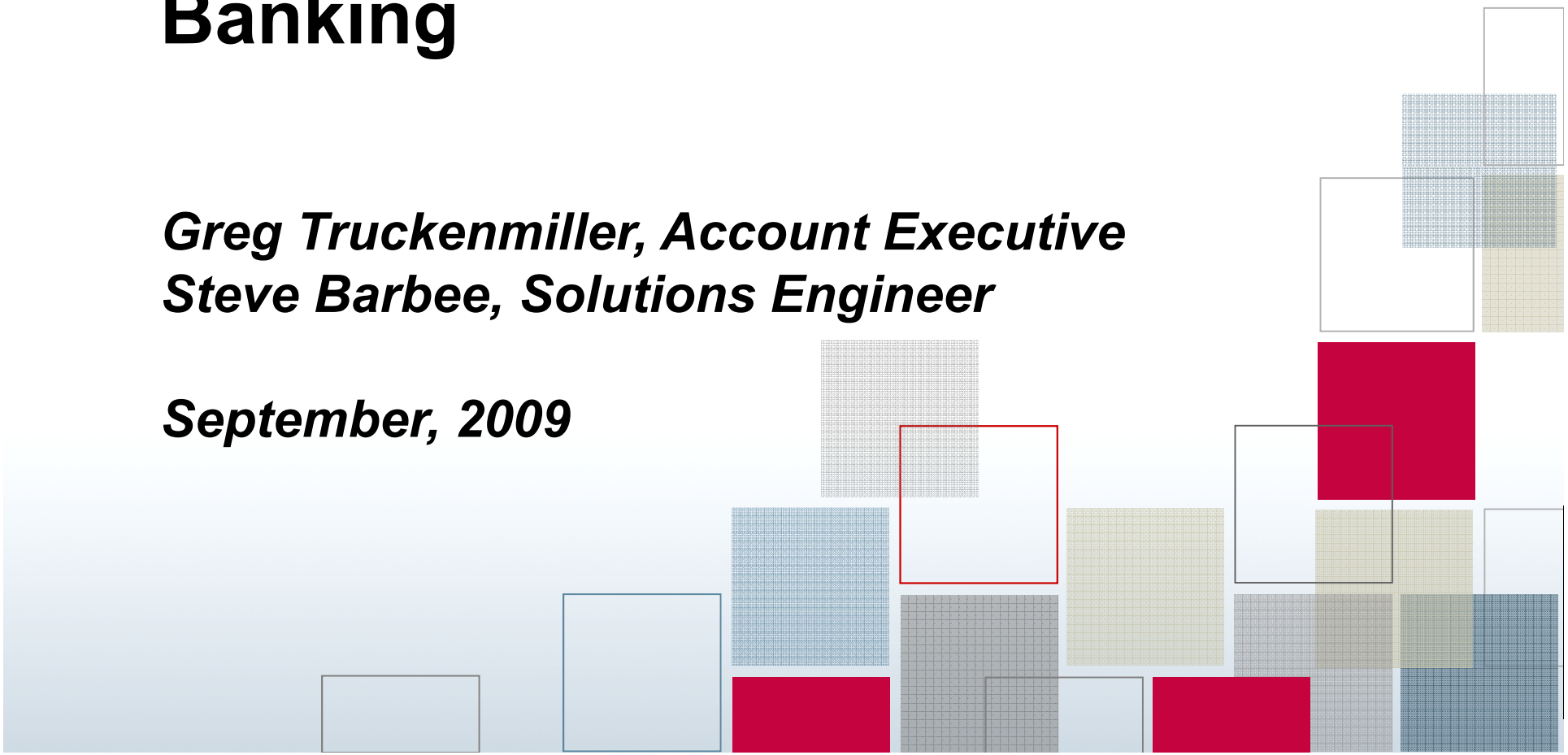
Data Mining Demonstration: PASW Modeler 13 for Retail Banking

Greg Truckenmiller, Account Executive
Steve Barbee, Solutions Engineer

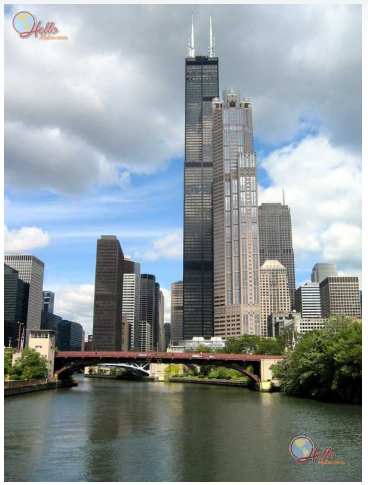
September, 2009



Agenda

- Welcome to SPSS!
- Bank on Success
- What is Data Mining & Predictive Analytics?
- PASW Modeler demonstration: 2 Scenarios
  - Art of the cross-sell
  - Who is in danger of attriting?
- Summary
- Q&A

SPSS Company Background



- **Software company**
  - Public-listed on NASDAQ – IBM to acquire
  - 40 year old company with heritage in analytic technologies
  - Top 25 software company
  - Operations in over 60 countries
- **Leadership**
  - Market leader in predictive analytics
  - Recognized as leader by *Forbes*, *BusinessWeek*, *Intelligent Enterprise*, *InfoWorld*, *CRM Magazine*, and others
- **Proven track record**
  - More than 95% of FORTUNE 1000 are SPSS customers
  - All of the top 10 global financial services companies (insurance, banking, brokerage)
  - 80% of the top 10 telecommunication services companies
  - 84% of the top 25 retailers worldwide
  - 96% of the top 25 global market research firms
  - More than 80% of the top consumer packaged goods companies

Leader in Predictive Analytics

"the strongest combination of features across the whole spectrum of elements required to define a world-class analytics suite. "

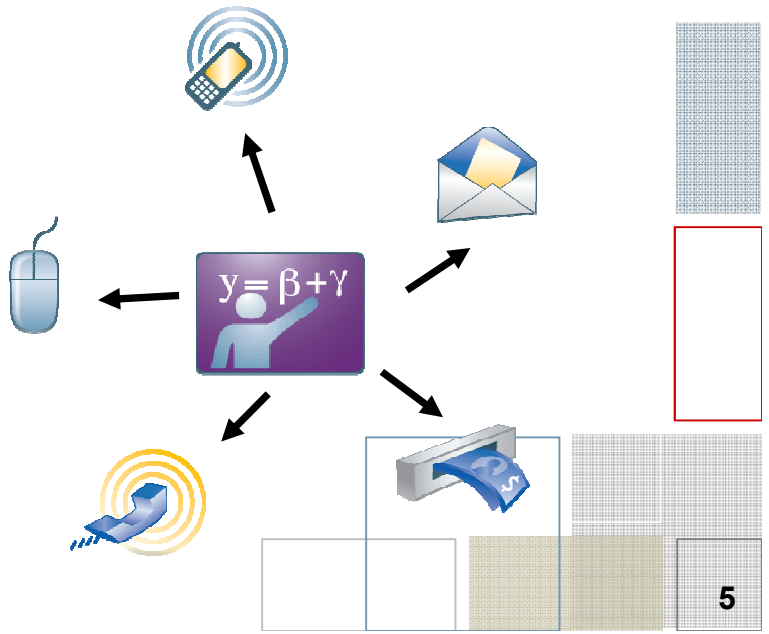
Leader – Magic Quadrant for Customer Data Mining, June 2008

"Some of the highest ROI scores Nucleus has ever seen"

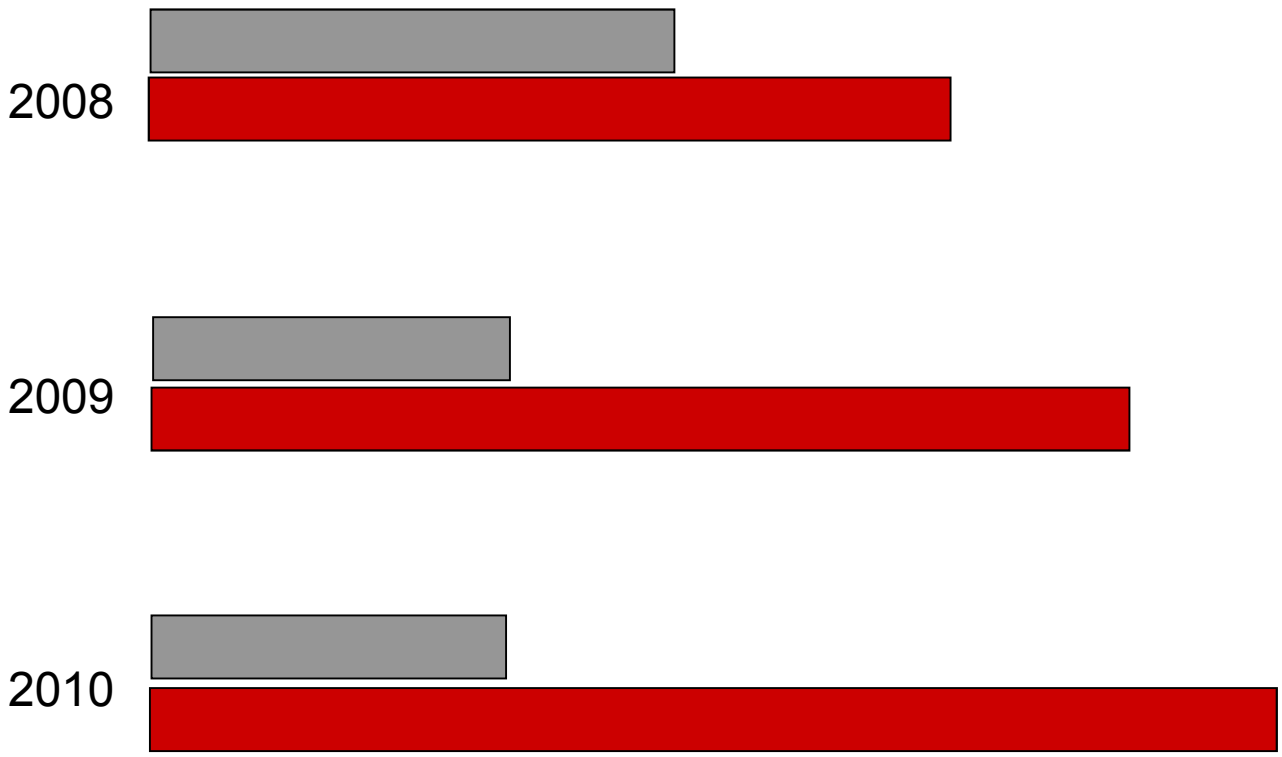
"Company to Watch" in business intelligence

Retail Banking: State of the Union
"Play to our strengths"

- Grow current customers vs. Acquire new
  - Product ownership penetration
  - Customer relationship profitability
- More meaningful customer interactions
  - One to one relationships
  - Interest me today with something I want
- Touch Points galore!
  - Direct mail
  - Online banking
  - Mobility banking
  - Inbound help centers
  - Outbound call campaigns
  - ATM's



Retail Banking: State of the Union "Do More With Less"



Retail Banking: State of the Union "The Solution"

- Use Your Data
  - Behavioral
  - Operational
  - Demographic
  - Text
- Use Your Channels
  - Mail
  - Web
  - Phone
  - Retail
- Realize Better Business Results
  - Lower mailing cost, higher offer response
  - Lower churn among profitable customers
  - Better cross-sell rates

The Right Prospect, the Right Offer

- Ambition:
  - Eliminate mailings to prospects least likely to respond
  - Migrate Product Managers from “gut feel” approach to data-based approach
- Less likely to “annoy” customers with unwanted offers
- Mined through past sales data to find best prospects for the product offering
- Easily explain predictive models to product managers

HSBC

- **380 retail branches in US**

- **33% fewer mail pieces**
- **Brought in 95% of the revenue of the previous campaign**
- **More nimble with timely offers due to short turnaround time**

Mid-Sized Bank Improves Customer Profitability and Product Ownership Rates

- Ambition:
  - Create a new premium, fee-based account to drive more profitable relationships
  - Must not cannibalize from existing customer base
- PASW Modeler found the *ideal* customer profile for the new checking account
- Considered data from transactions (ATM visits, NSF, etc.), demographics (Claritas) and product ownership
- Microtargeted offers to 6% of the customer base

- **100,000 customers**
- **55 retail branches**

- **10% campaign response-rate**
- **Minimized cannibalization within current customer base**

Cablecom
Improves Retention of Profitable Customers

Background
- Swiss Telco and TV operator
- Over 54% of Swiss television households receive cable service from Cablecom

Business Goals
- Reduce the number of profitable customers who defect to competitors

Solution
- Uses PASW Modeler to profile likely defectors
- Added text mining to data mining through PASW Text Analytics to better understand the reason for leaving

Results
- **Reduce churn from 19% to 2%**
- **Improved effectiveness of predictive models**
- **Now keep retention rates high**
- **51% of Cablecom dissatisfied customers became company promoters (very satisfied customers) after two months**

The Power of Predictive Marketing

	Without Data Mining	*With* Data Mining
Response Rate	0.5%	2.5%
	Another campaign? Better creative?	
	Revenue Cost	Revenue Cost
Responders		
Solicited		

What is Data Mining?

- "…the exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules" -- Berry & Linoff*
- "**Predictive analytics** is a set of business intelligence technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events." -- TDWI Research**

* From Data Mining Techniques: For Marketing, Sales & Customer Support, Michael J.A. Berry & Gordon LInoff, p.5

** "Predictive Analytics," What Works in Data Integration, TDWI Research, Vol.23, 2007, p.49

What is Data Mining?

Table Edit Generate Help

id	age	gender	responder
ID12106	57	FEMALE	YES
ID12108	58	MALE	NO
ID12120	31	MALE	NO
ID12138	36	FEMALE	YES

Discovering meaningful patterns in your data

What is Data Mining?

As the data grows…

Table Edit Generate Help

id	age	gender	responder
ID12106	57	FEMALE	YES
ID12108	58	MALE	NO
ID12120	31	MALE	NO
ID12138	36	FEMALE	YES

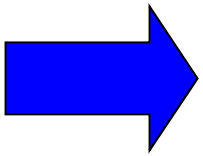


Table Edit Generate Help

id	age	gender	responder
ID12106	57	FEMALE	YES
ID12108	58	MALE	NO
ID12120	31	MALE	NO
ID12121	61	MALE	YES
ID12138	36	FEMALE	YES

The relationships become more complicated

What is Data Mining?

Now we have 8 additional inputs….. against our target

Target

Table Edit Generate Help

Inputs

id	age	gender	region	income	married	children	car	save_act	current_act	mortgage	responder
ID12101	48	FEMALE	INNER_CITY	17546.0	NO	1	NO	NO	NO	NO	YES
ID12102	40	MALE	TOWN	30085.1	YES	3	YES	NO	YES	YES	NO
ID12103	51	FEMALE	INNER_CITY	16575.4	YES	0	YES	YES	YES	NO	NO
ID12104	23	FEMALE	TOWN	20375.4	YES	3	NO	NO	YES	NO	NO
ID12105	57	FEMALE	RURAL	50576.3	YES	0	NO	YES	NO	NO	NO
ID12106	57	FEMALE	TOWN	37869.6	YES	2	NO	YES	YES	NO	YES
ID12107	22	MALE	RURAL	8877.07	NO	0	NO	NO	YES	NO	YES
ID12108	58	MALE	TOWN	24946.6	YES	0	YES	YES	YES	NO	NO
ID12109	37	FEMALE	SUBURBAN	25304.3	YES	2	YES	NO	NO	NO	NO
ID12110	54	MALE	TOWN	24212.1	YES	2	YES	YES	YES	NO	NO

The data mining methods discovers meaningful patterns in your complex data

The goal is to find actionable insights

Business intelligence compared

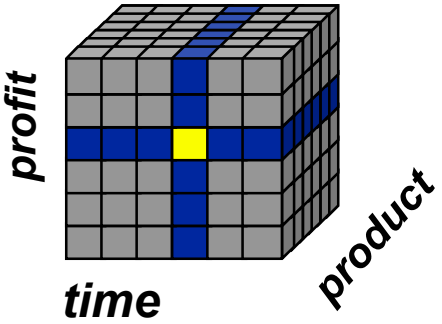
Query/Reporting

- Hypothesis-driven
- Manual

'What were sales of product X in October'

OLAP

- Hypothesis-driven
- Manual



'Drill down October Sales of product X at 4% profit level, all regions'

Data Mining

- Data- & Goal-driven
- Creates Hypotheses
- Automatic

Goal = 'apply for card':

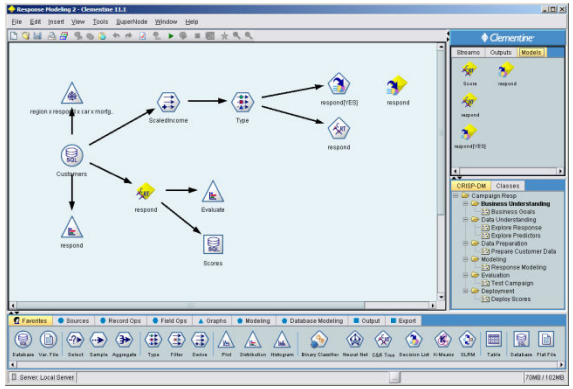
'If period = week 40-48
and region = NW
and branch = S
then apply for card =
'denied'

Reports & Graphs

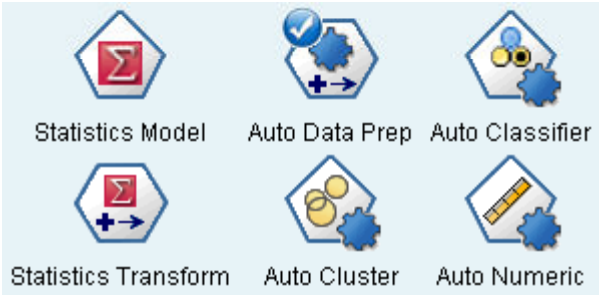
Executable Decision Model

PASW Modeler Capabilities

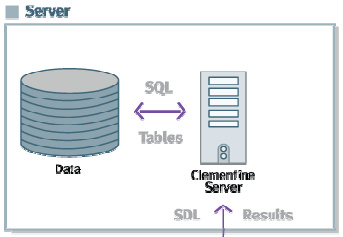
- Easy to Learn / Visual Design Paradigm
  - Visual approach - no writing code!
  - Comprehensive range of data mining methods
  - Rexer Analytics Survey (& kdnuggets.com) – *Most popular workbench*



- Powerful Automated modeling
  - Automatically prepares data
  - Automatically finds the best mining model
  - Mines text, web & survey data
  - Integrates with PASW Statistics



- Open & Scalable architecture
  - No proprietary database required
  - Easy ODBC connection to most databases
  - Leverage your existing IT investment
  - Scales to meet global demand and enterprise volume



Demo Scenarios

Grow: Cross-Sell

- Which customers should I include in campaigns?
  - Pension, Savings, Car loan, Mortgage

Retention

- Which valuable customers should we campaign to retain?

Grow

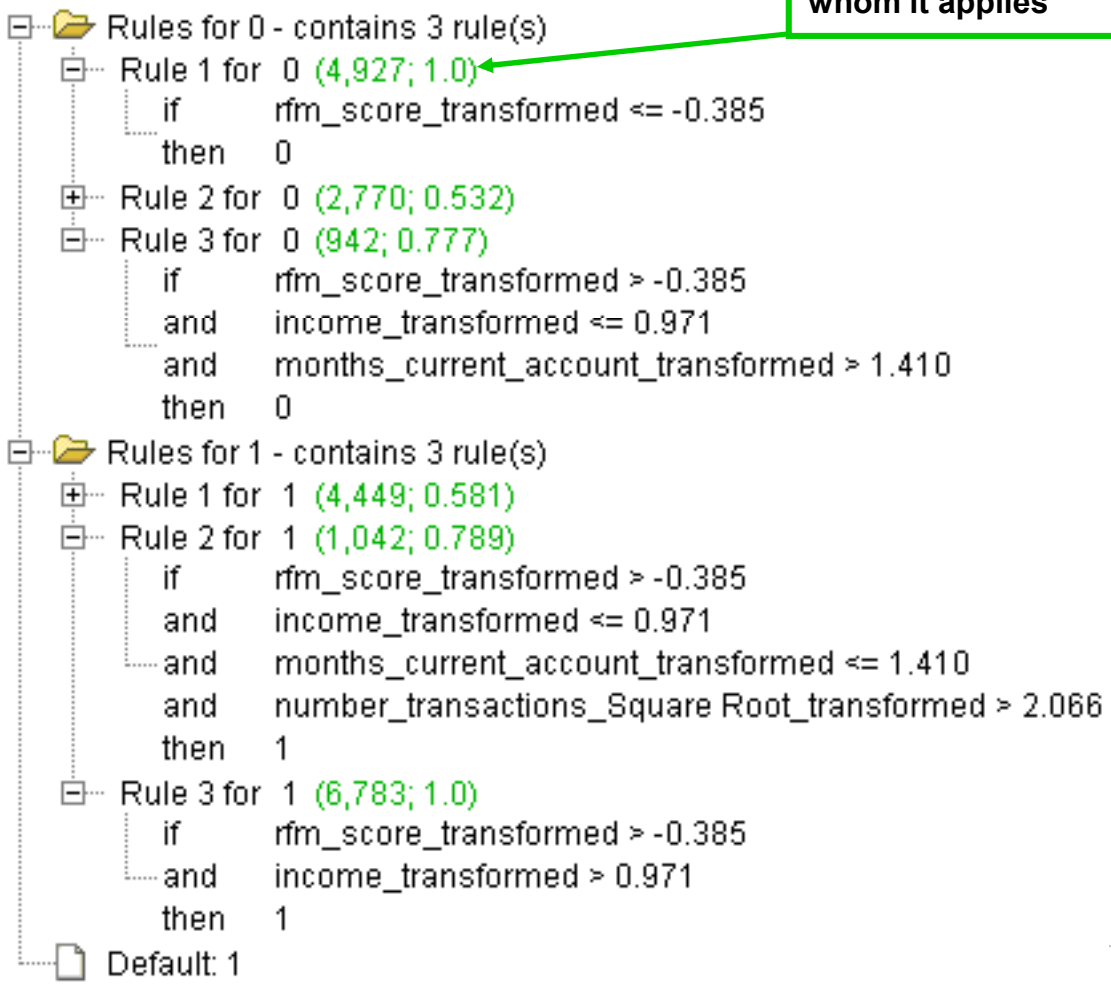
- Cross-Sell – what is the next most likely product?
  - For customer 64199, the next likely purchase is a car loan with a probability of 96.8%

	customer_id	campaign	response	1st Likely	Propensity-1	2nd Likely	Propensity-2	3rd Likely	Propensity-3
1	64199	Mortgage	1	Car loan	0.968	Savings	0.957	Pension	0.868
2	60877	Car loan	1	Savings	0.957	Mortgage	0.893	Pension	0.868
3	24535	Car loan	1	Savings	0.957	Mortgage	0.893	Pension	0.868
4	67764	Car loan	1	Savings	0.957	Mortgage	0.893	Pension	0.868
5	67345	Car loan	1	Savings	0.957	Mortgage	0.893	Pension	0.868
6	32010	Car loan	1	Savings	0.957	Mortgage	0.893	Pension	0.868
7	18523	Car loan	1	Savings	0.957	Mortgage	0.893	Pension	0.868
8	8521	Car loan	1	Savings	0.957	Mortgage	0.893	Pension	0.868
9	68741	Car loan	1	Savings	0.957	Mortgage	0.892	Pension	0.868
10	9274	Car loan	1	Savings	0.957	Mortgage	0.892	Pension	0.868
11	2423	Car loan	1	Savings	0.957	Mortgage	0.892	Pension	0.868
12	15833	Car loan	1	Savings	0.957	Mortgage	0.891	Pension	0.868
13	3445	Car loan	1	Savings	0.957	Mortgage	0.891	Pension	0.868
14	40395	Car loan	1	Savings	0.957	Mortgage	0.891	Pension	0.868
15	32719	Car loan	1	Savings	0.957	Mortgage	0.891	Pension	0.868
16	7759	Car loan	1	Savings	0.957	Mortgage	0.891	Pension	0.868
17	34890	Car loan	1	Savings	0.957	Mortgage	0.890	Pension	0.868
18	23642	Mortgage	1	Savings	0.957	Car loan	0.889	Pension	0.868
19	28385	Car loan	1	Savings	0.957	Mortgage	0.889	Pension	0.868
20	2248	Car loan	1	Savings	0.957	Mortgage	0.889	Pension	0.868

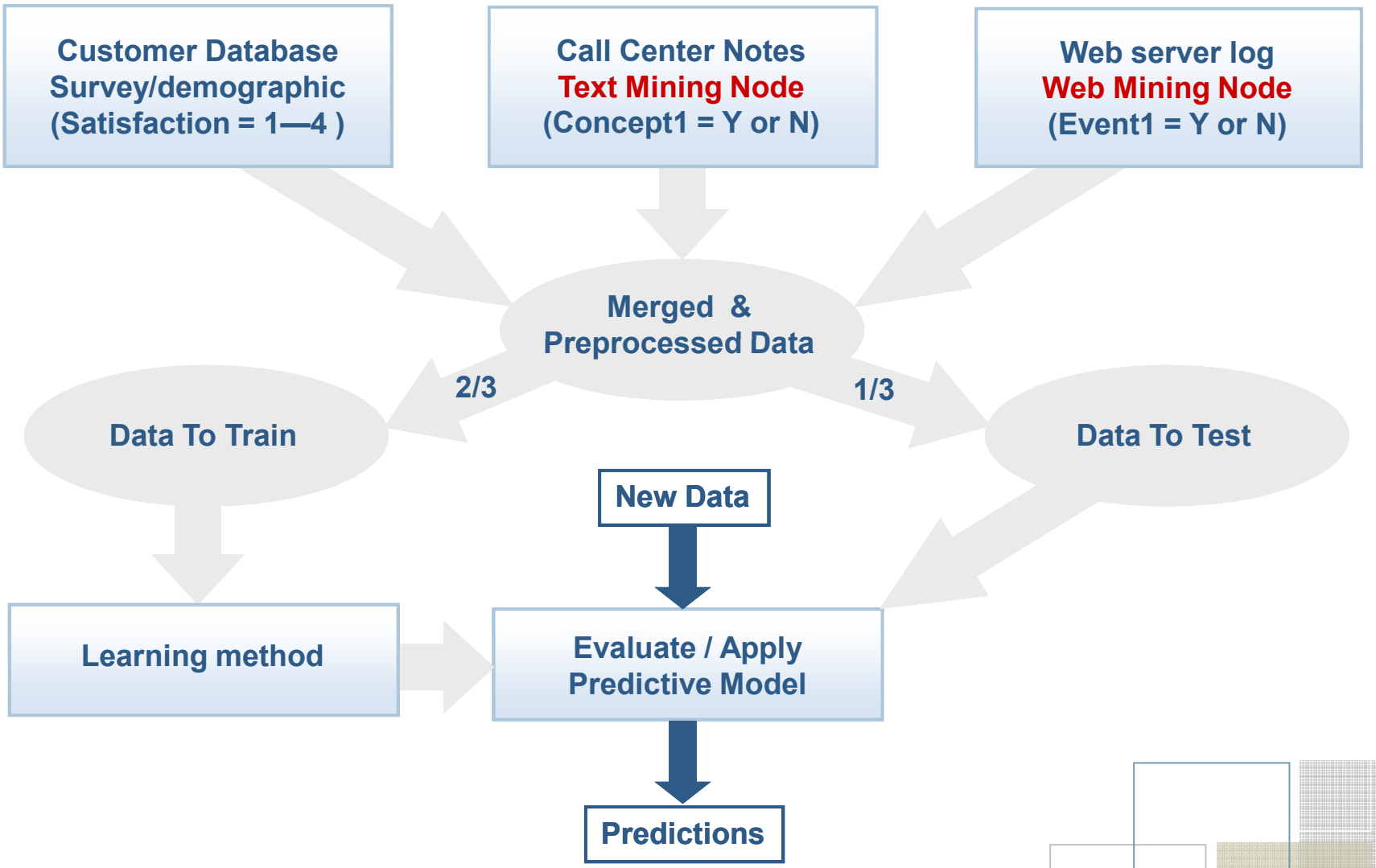
Car Loan Campaign Response Model

- Some C & R Tree Rules

This rule is 100% accurate for the 4,927 customers to whom it applies

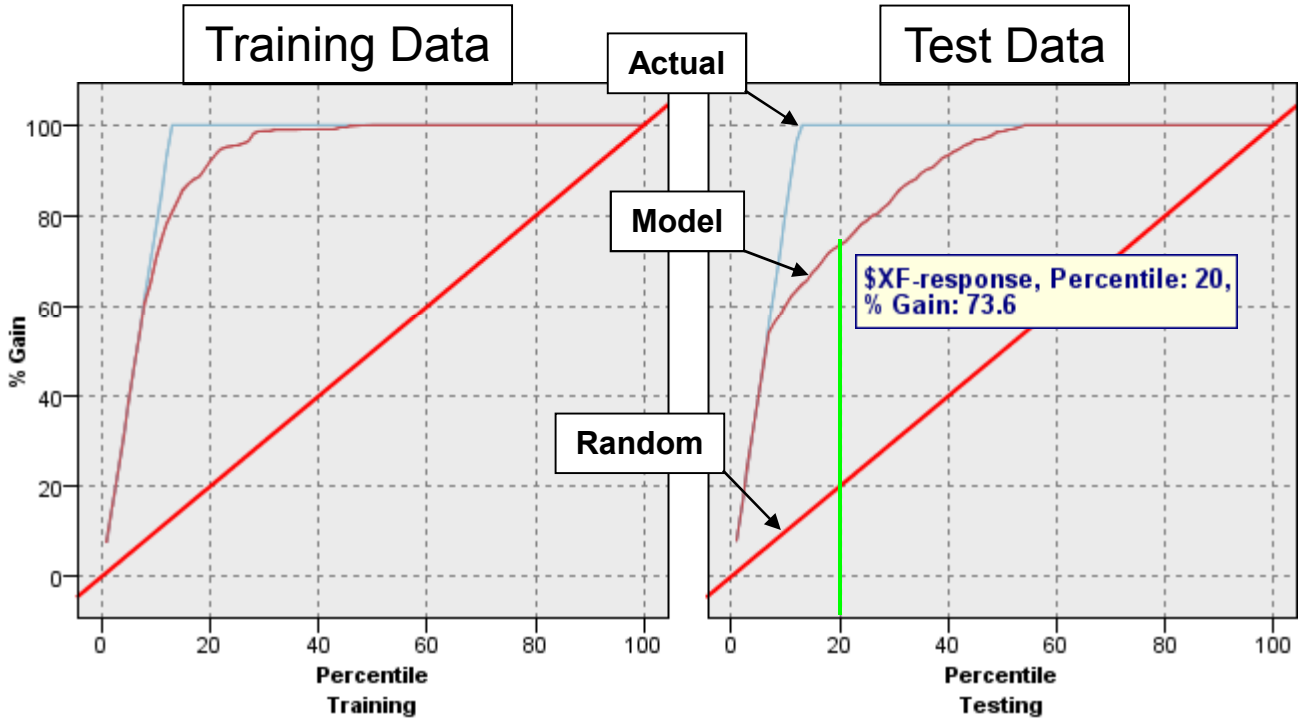


How a Predictive Model is Made & Tested



Campaign Model Evaluation

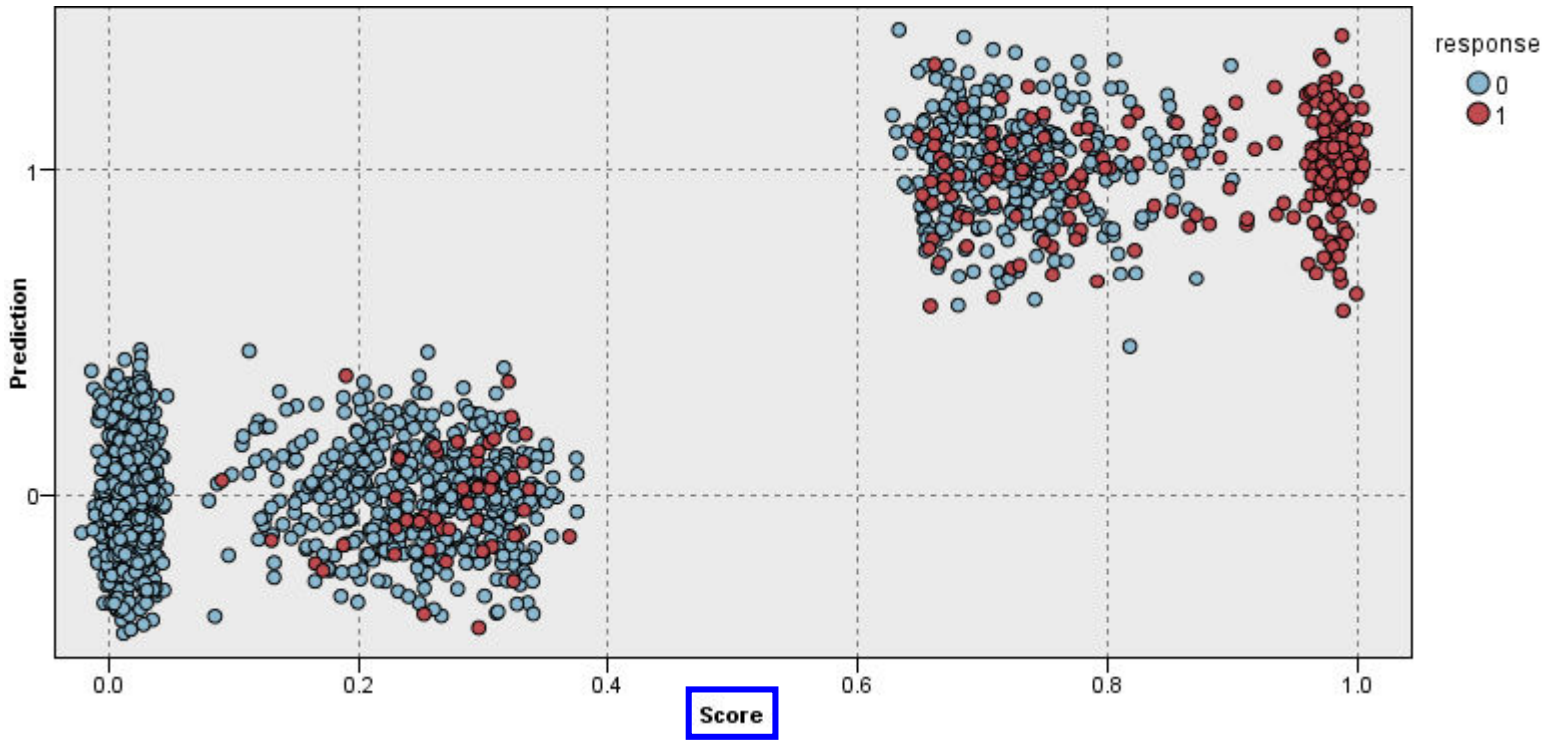
- Top 51% hits 77.8% of “Leaving” customers



'Partition'	1_Training		2_Testing	
Correct	12,821	84.87%	5,950	80.85%
Wrong	2,285	15.13%	1,409	19.15%
Total	15,106		7,359	

Deploy Your Predictions for Action

- Those Likely to Respond To a Car Loan Campaign have a high Score



Demo Scenarios

Cross-Sell

- Which customers should I include in campaigns?
  - Pension annuity, CDs, Car loan, Mortgage

Retention

- Which valuable customers should we campaign to retain?

Retention Model Rules

- From C&R Tree Portion of Model

Rules for Leaving

Rules for Leaving - contains 3 rule(s)

- Rule 1 for Leaving (96; 0.667)
  - if Number of Cards <= 2.500
  - and Income > 31879.500
  - and Loan Amount <= 10,700
  - and Income <= 36,209
  - then Leaving
- Rule 2 for Leaving (874; 0.965)
  - if Number of Cards > 2.500
  - and Marital Status in ["Married"]
  - then Leaving
- Rule 3 for Leaving (1,013; 0.668)
  - if Number of Cards > 2.500
  - and Number of Cards > 4
  - and Income <= 28,924
  - and Marital Status in ["Single" "Divorced"]
  - then Leaving

Rules for Staying

Rules for Staying - contains 5 rule(s)

- Rule 1 for Staying (1,539; 0.758)
  - if Number of Cards <= 2.500
  - and Income <= 31879.500
  - then Staying
- Rule 2 for Staying (399; 0.594)
  - if Number of Cards <= 2.500
  - and Income > 31879.500
  - and Loan Amount <= 10,700
  - and Income > 36,209
  - then Staying
- Rule 3 for Staying (119; 0.739)
  - if Number of Cards <= 2.500
  - and Income > 31879.500
  - and Loan Amount > 10,700
  - then Staying
- Rule 4 for Staying (210; 0.843)
  - if Number of Cards > 2.500
  - and Number of Cards <= 4
  - and Marital Status in ["Single" "Divorced"]
  - then Staying
- Rule 5 for Staying (94; 0.553)
  - if Number of Cards > 2.500
  - and Number of Cards > 4
  - and Income > 28,924
  - and Marital Status in ["Single" "Divorced"]
  - then Staying

Retention Model Comparison

- Without Demographic Data, Accuracy falls from 78.6% to 75.6% for “Leaving” customers

Individual Models

Comparing All Data-Retention Risk with Retention Risk

'Partition'	1_Training		2_Testing	
Correct	3,273	75.35%	1,076	78.65%
Wrong	1,071	24.65%	292	21.35%
Total	4,344		1,368	

Coincidence Matrix for All Data-Retention Risk (rows show actuals)

'Partition' = 1_Training	Leaving	Staying
Leaving	1,577	648
Staying	423	1,696
'Partition' = 2_Testing	**Leaving**	**Staying**
Leaving	187	89
Staying	203	889

Comparing No Demographic-Retention Risk with Retention Risk

'Partition'	1_Training		2_Testing	
Correct	3,198	73.62%	1,035	75.66%
Wrong	1,146	26.38%	333	24.34%
Total	4,344		1,368	

Coincidence Matrix for No Demographic-Retention Risk (rows show actuals)

'Partition' = 1_Training	Leaving	Staying
Leaving	1,579	646
Staying	500	1,619
'Partition' = 2_Testing	**Leaving**	**Staying**
Leaving	188	88
Staying	245 →	847

Predicted to Leave

Actually Staying

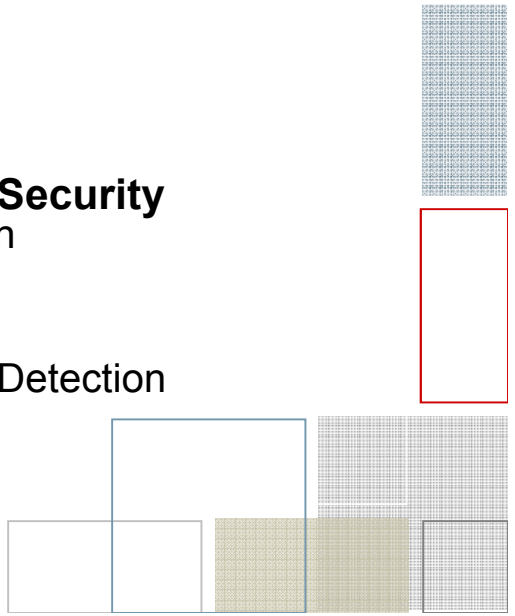
Fewer False Positives means less mailing cost and wasted incentives on those intending to stay!

PASW Key Capabilities Review

- Data Mining Workbench
  - **1st to introduce** intuitive, visual approach
  - **No need to program** lines of code
  - Consistently **most-popular** data mining workbench
  - **Scalable**, enterprise-level capabilities
  - **Push** results out to touch points
  - **Automatic methods** to make data mining easy
- Many case studies and examples of high-leverage applications in retail banking
  - Growth through cross-selling
  - Retention
  - Product development
  - Risk
  - Fraud

Who Else is Using Data Mining?

- **Direct Marketing**
  - Response modeling, cross-selling, segmentation
- **Retail**
  - Pricing
  - Shelf-space Optimization
  - Offer Optimization
- **Telecommunications**
  - Churn reduction, fraud detection, cross-sell
- **Banking**
  - Credit scoring, loan approval, fraud detection
- **State and Federal Gov't**
  - Tax compliance, Fraud detection
- **Insurance**
  - Fraud detection
  - Customer acquisition
- **Non-profit Companies**
  - Donations, Major Gifts, Advocacy
- **Stock Market**
  - Stock Timing, Fund Management, Portfolio Optimization
- **Pharmaceuticals**
  - Drug discovery, Combinatorial Chemistry
- **MLB, NBA and NHL**
  - Optimizing rosters, Scoring Tendencies
- **Engineering**
  - Signal Detection
  - Fault Diagnostics
- **Computer Systems/Security**
  - Intrusion Detection
- **Defense**
  - Automatic Target Detection
  - Missile Guidance



Thank You!
Q&A

Greg Truckenmiller, Account Executive
312-651-3254
gtruckenmiller@spss.com

Steve Barbee, Solutions Engineer