

# Approaches to Incremental Response Modeling

**Doug Freud**  
Technical Director  
[doug.freud@kxen.com](mailto:doug.freud@kxen.com)

**Robert Cooley**  
CTO Optimine Software  
[rob.cooley@optiminesoftware.com](mailto:rob.cooley@optiminesoftware.com)



September 14, 2009 THE DATA MINING AUTOMATION COMPANY™

# Rules for Retail Marketing



# Agenda

---

- **KXEN Background**
- **KXEN Overview & SAS**
- **Quick KXEN Demo**
- **Approaches to Incremental Response Modeling**
- **Case Studies**
- **Q&A**
- **Lunch**



# KXEN , Who we are



## KXEN - Company Background

- **Founded in July 1998, Delaware corporation**
- **HQ in San Francisco, R&D in Paris, Operations in US, UK, France – Exclusive Distributors throughout the World**
- **Investors:**
  - Sofinnova US & France, Saints Capital, XAnge, Motorola Ventures
- **Auditors : KPMG, US GAAP compliant**
- **CAGR above 30% since 2004**
- **KXEN Analytic Framework in 5th major release**
  - Available on Windows, Linux, Solaris, HP-UX, AIX,
  - Available in French, English, Japanese, Korean, Chinese, Russian



## KXEN & SAS Value Proposition

- **We are not a replacement for SAS**
- **Most of our Customers are also SAS customers**
- **SAS is typically used for:**
  - Building Analytical Data Sets
  - Data Quality
  - Data Exploration
  - Scoring
  - Evaluation of Campaign effectiveness

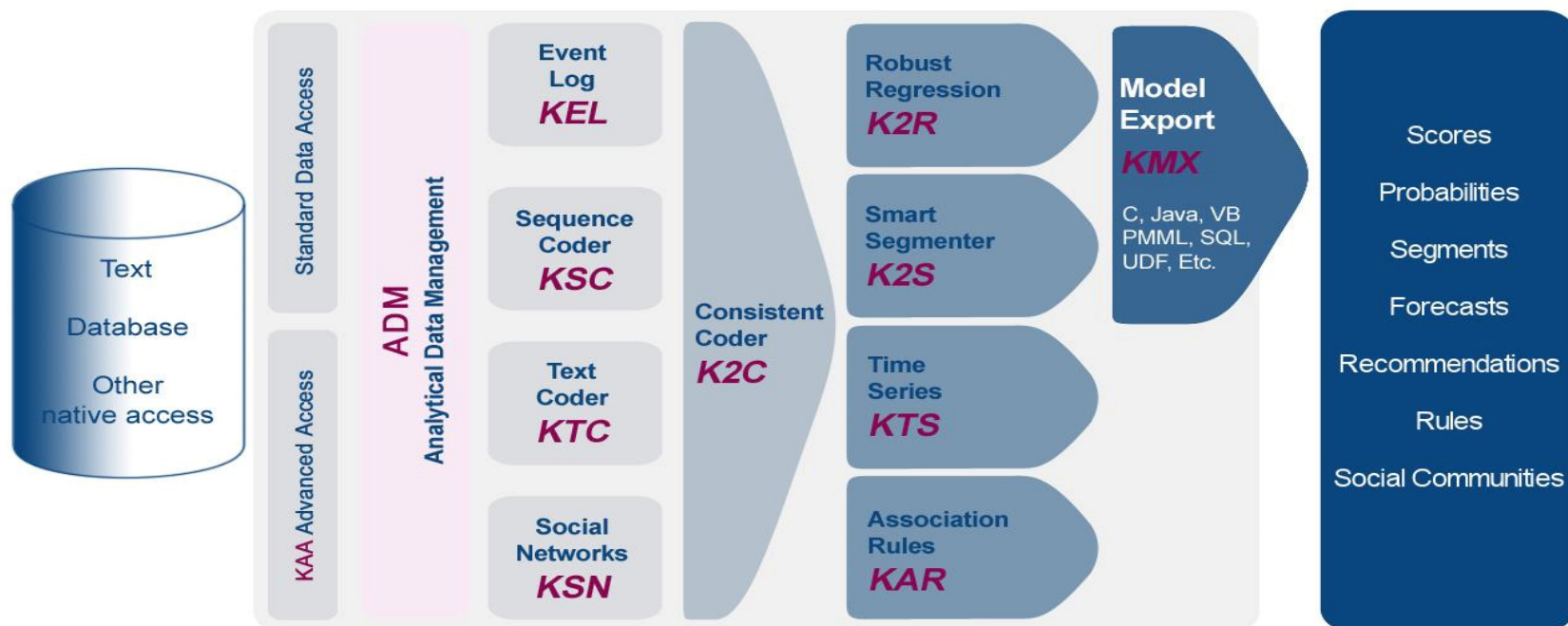


## KXEN Use Cases for SAS Customers

- **Need 100's or 1000's of models**
- **Need to frequently rebuild models**
- **Need to automate building of models**
- **Automate Variable selection with 100's or 1000's of variables**



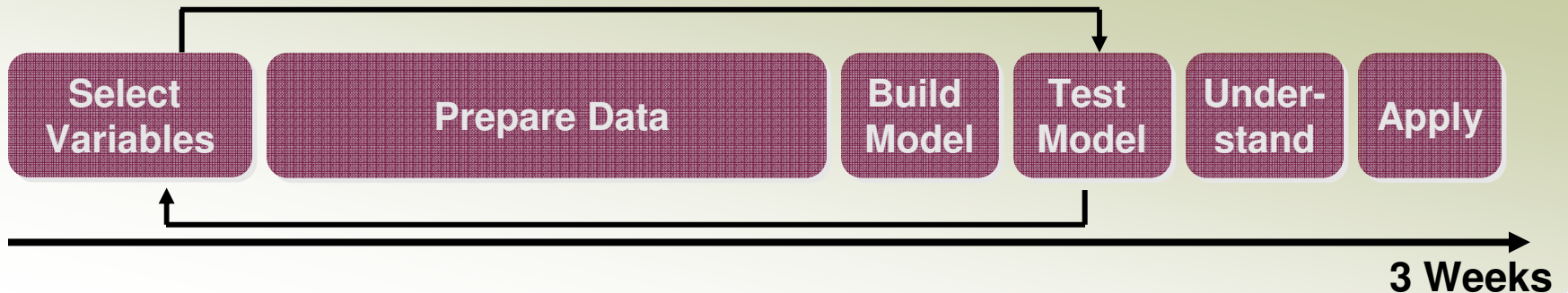
# The KXEN Analytical Framework (V5.1)



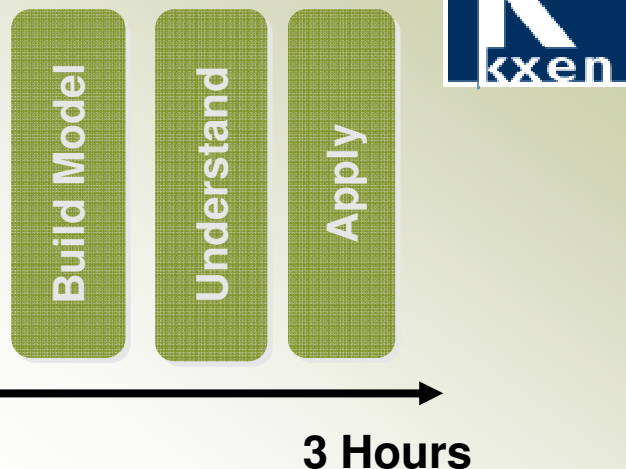
- Unique Data Fusion approach
- Built upon industry standards like Java Data Mining (JDM), Web-services, Predictive Modeling Markup Language (PMML), SQL.
- Provides SSL and PAM authentication and activity logging
- Support for Unicode and 64-bit architectures.

## Do More...Modeling Automation is Key!

### Legacy Data Mining



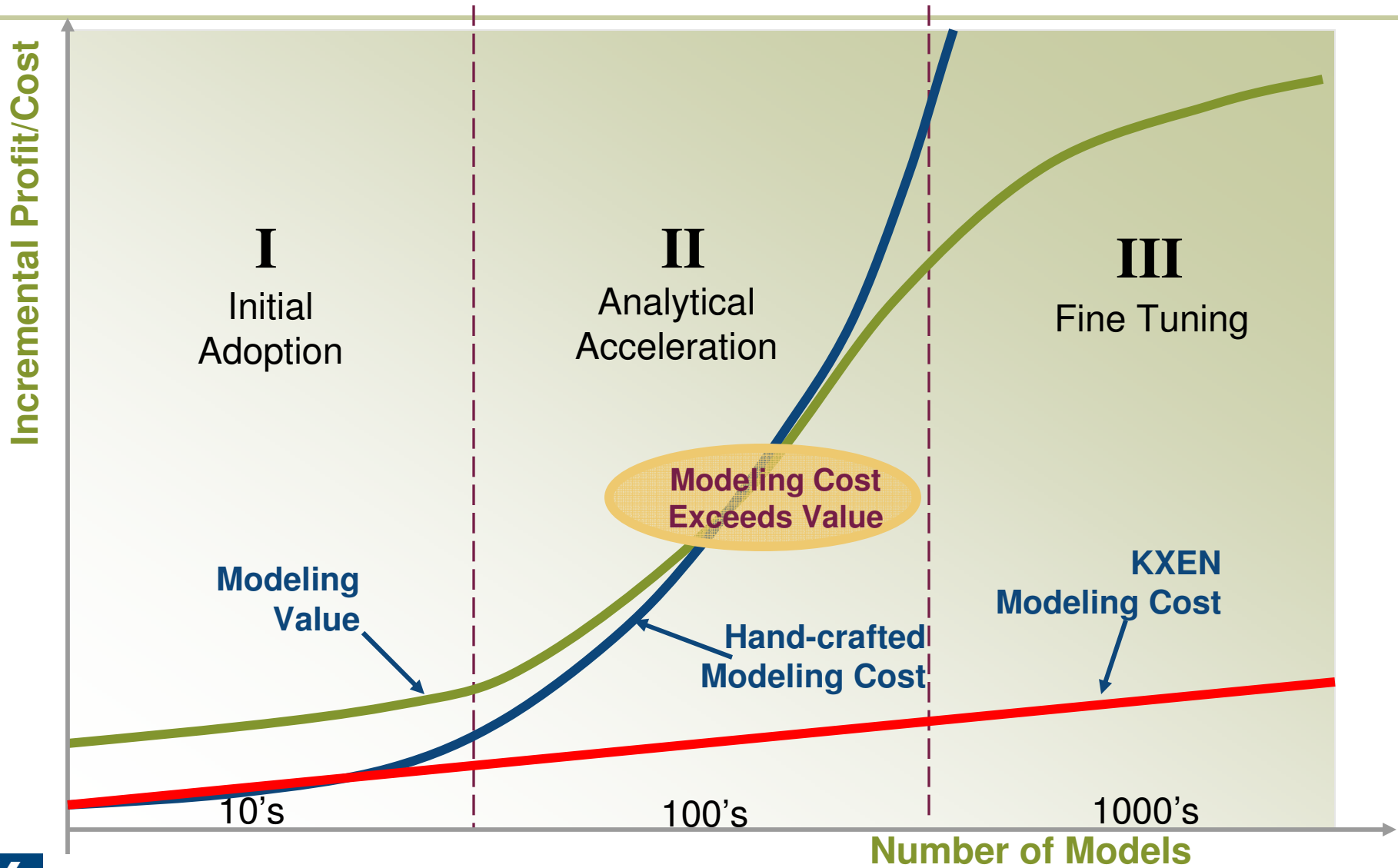
### Next Generation



### KXEN Automates

- ✓ Variable selection
- ✓ Data preparation
  - Variable encoding
  - Missing value handling
  - Outlier handling
  - Binning and banding
- ✓ Model Fitting
- ✓ Model Testing
- ✓ Model Reporting
- ✓ Model Deployment

# The Productivity Bottleneck



## The Problem



### Launch New Campaign

- To drive people into the store, web site, call center, ...

### Incentive Offered

- % Off
- \$ Off
- BOGO (by one get one)

**Response Rates for the campaign were very high but...**

# The Problem

## Poor Results

- Overall sales are barely affected or drop
- Margin actually decreases in the weeks after the campaign!
- ROI is low or potentially negative!
- Churn/attrition increases

## Because..

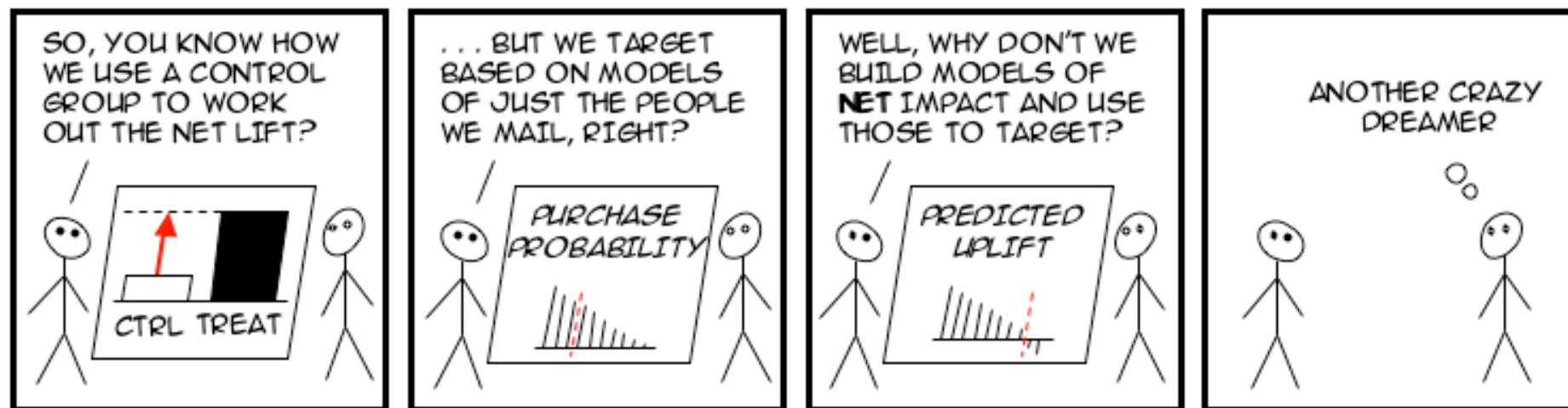
- We incentivized the wrong customers to purchase
- We sent offers to customers that may have purchased anyway!
- We may have cannibalized future revenues



## What Happened?

### Traditional Response Models

- “Find” customers most likely to exhibit a certain behavior.
- Unable to distinguish between, “The Persuadables,” and “The Sure Things.”



[HTTP://SCIENTIFICMARKETER.COM](http://SCIENTIFICMARKETER.COM)

COPYRIGHT © NICHOLAS J RADCLIFFE 2007. ALL RIGHTS RESERVED.



## Customer Segments

- **The Persuadables** : customers who only respond to the marketing action because they were targeted
  - Only segment that provides incremental positive response
- **The Sure Things** : customers who would have responded whether they were targeted or not
  - Campaigning to this segment, will have adverse effect on revenue/profit
- **The Lost Causes** : customers who will not respond irrespective of whether they are targeted
  - Irrelevant
- **The Do Not Disturbs** : customers who are less likely to respond because they were targeted
  - Campaigning to this segment, may cause churn/attrition/cancellation

## The Solution – Find the Persuadables



- **“The Persuadables” will typically produce the highest value**
  - Value may be defined as response, ROI, margin, spend, lifetime value....
- **Removes the negative effects from retention campaigns**
  - Distinguish and exclude “The Do Not Disturbs” from the campaign

## Why is this is a Hard Problem?



### Technical

- Time to develop Incremental Lift models is longer
  - Requires multiple models
  - Models must be converted to provide comparable probabilities
- Refresh rates are often more frequent
- High cost per models means that prototyping various options becomes too expensive

### Design

- Requires a large enough control group to ascertain the pool of responders given no treatment.
- Marketing may want to mail to every potential responder
- The proper expectations must be set by the business. Higher ROI but...
  - Reduced mail volumes
  - Reduced response rates

# Basic Approach for Incremental Response

## Approach

- **Build two models:**
  - One on the Mailed population
  - One on the Control (Not Mailed) population
- **At the time of scoring, score each customer with both models, creating 3 probabilities:**

*Probability of Incremental Response =  
(Probability of response if mailed -  
Probability of response if not mailed)*
- **Mail to the customers with the highest incremental response**



# Basic Approach for Incremental Response

Customer	Scored by Mail Model $P(\text{response})_{\text{mailed}}$	Scored by Control model $P(\text{response})_{\text{not mailed}}$	Score Difference (Incremental Response)
1	0.7536	0.7502	0.0034
2	0.5824	0.4795	0.1029
3	0.0579	0.0285	0.0294
4	0.3754	0.3841	-0.0087
5	0.0102	0.0101	0.0001
6	0.6421	0.6954	-0.0533
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
1000	0.1873	0.1654	0.0219



Customer	Incremental response probability
2	0.1029
.	.
.	.
3	0.0294
.	.
1000	0.0219
.	.
.	.
1	0.0034
.	.
5	0.0001
.	.
4	-0.0087
.	.
6	-0.0533

**Sort by descending score difference**



**Select from top to bottom for targeting**

# Basic Approach for Incremental Spend

## Problem

- **Customers that would have shopped anyway may spend MORE if provided an incentive**
- **Some Customers will only shop with the incentive, yielding a net loss**

## Approach

- **Build 4 models to develop Incremental Spend**
  - **(Probability of response if mailed \* Expected revenue if mailed and responded) – (Probability of response if not mailed \* Expected revenue if not mailed and responded)**
- **Sort from highest to lowest**
- **May want to consider predicting margin rather than spend**



## Approach for Incremental Lifetime Value

### Problem

- Do customers that get offers come back with a higher frequency?
- Some customers stockpile – profit today yields loss tomorrow

### Approach

- Develop models that predict lifetime value over a window of time if they are given an offer and if they aren't given the offer.
  - Expected value in week1 +  
Expected value in week2 +  
Expected value in week3...



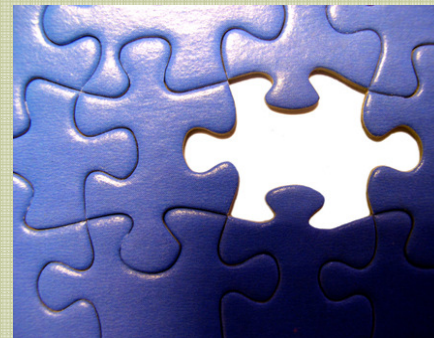
# How is KXEN Uniquely Suited for these Modeling Approaches? (1)

## Automatic Data Preparation Means that...

- The modeling effort becomes “iterations on a theme”
  - Manipulate data once, split it into several groups
  - Model automatically – for each group
  - Try a number of approaches to determine the one that performs the best on the validation set
- Cost per model ~\$500 means that more models can be built for each region, rather than 1 generic model

## Automatic Generation of True Probabilities

- Historically a score would have had to have been converted into a probability
- No need to balance data prior to modeling. KXEN handles response rates from 1% to 50%



# How is KXEN Uniquely Suited for these Modeling Approaches? (2)

## The KXEN Incremental Modeling:

### Components

- KXEN Software
  - Classification
  - Regression
  - Segmentation
  - Data Preparation
  - Analytic Dataset Creation
- Model Management Tools
  - Scripts
- Implementation Services
- Implementation Guide
  - Best practice methodologies
  - Test and Control population guidance

### Methodologies Include

- **Probability**
- **Spend**
- **Probability \* Spend**
- **Weighted Probability**
- **Top Decile Probability**
- **Two Stage**
  - Probability threshold followed by Spend
- **Micro Segmentation**
  - Different set of models for each segment
  - One set of scores depending on segment assignment
- **Category Rollup**
  - Different set of models for each category
  - Sum of scores for each category

# Lowes: 700% Increase in Productivity

## Existing Practice

- Campaigns require models to be scored against 20 million customers
- Roughly 5 models per analyst per year with 1<sup>st</sup> generation data mining workbench
- About 100 attributes considered for each modeling project
- May be sending offers to customers that would have responded anyway

## Goal

- Improve profit, not just per campaign revenue
- Target more campaigns
- Utilize existing modeling staff and infrastructure



## Lowes: 700% Increase in Productivity

### Solution

- Use KXEN's Data Mining Automation Solution for model training instead of 1<sup>st</sup> generation data mining workbench
- Push SQL scoring code back into Teradata warehouse
- Use Incremental Modeling to improve corporate profitability

### Results

- **700% increase in modeler productivity**
  - 35 – 40 models per year per analyst (up from 5)
- **1100 attributes considered for each modeling project (up from 100)**
- **Increased campaign response rates (>200% increase)**



THE DATA MINING AUTOMATION COMPANY™



# Large US Retailer: 250% Improvement in Campaign Profit

## Existing Practice

- **Roughly 200 direct mail campaigns run each year**
- **About 50 campaigns are targeted with “incremental profit” models**
  - Staff of 5 analysts
  - 2 weeks for model training and validation
  - 1 week for scoring/deployment
- **Remaining 150 campaigns are targeted with “business rules”**

## Goal

- **Improve profit (total margin on items purchased)**
- **Target all 200 campaigns with incremental profit models**
- **Utilize existing modeling staff and infrastructure**

## **Large US Retailer: 250% Improvement in Campaign Profit**

### **Solution**

- **Use KXEN Analytic Framework for model training instead of traditional data mining workbench**
- **Push SQL scoring code back into Teradata warehouse**

### **Results**

- **80% Reduction in Modeling and Deployment Time**
  - **2 days for model training/validation**
  - **1 day for scoring/deployment**
- **Targeted all 200 campaigns with incremental profit models**
- **150% Improvement in overall response rate**
- **250% improvement in profit**



## Next Steps

### For More Information

Contact a KXEN Representative to arrange a product demonstration and discussion about whether your organization can benefit from Incremental Response Modeling.

[sales-us@kxen.com](mailto:sales-us@kxen.com)

### To Contact our Speaker

Technical Director, KXEN  
[doug.freud@kxen.com](mailto:doug.freud@kxen.com)  
(630) 723-1744

