

Data Mining in Practice

Boston Area SAS Users Group

Michael J. A. Berry

Data Miners, Inc.

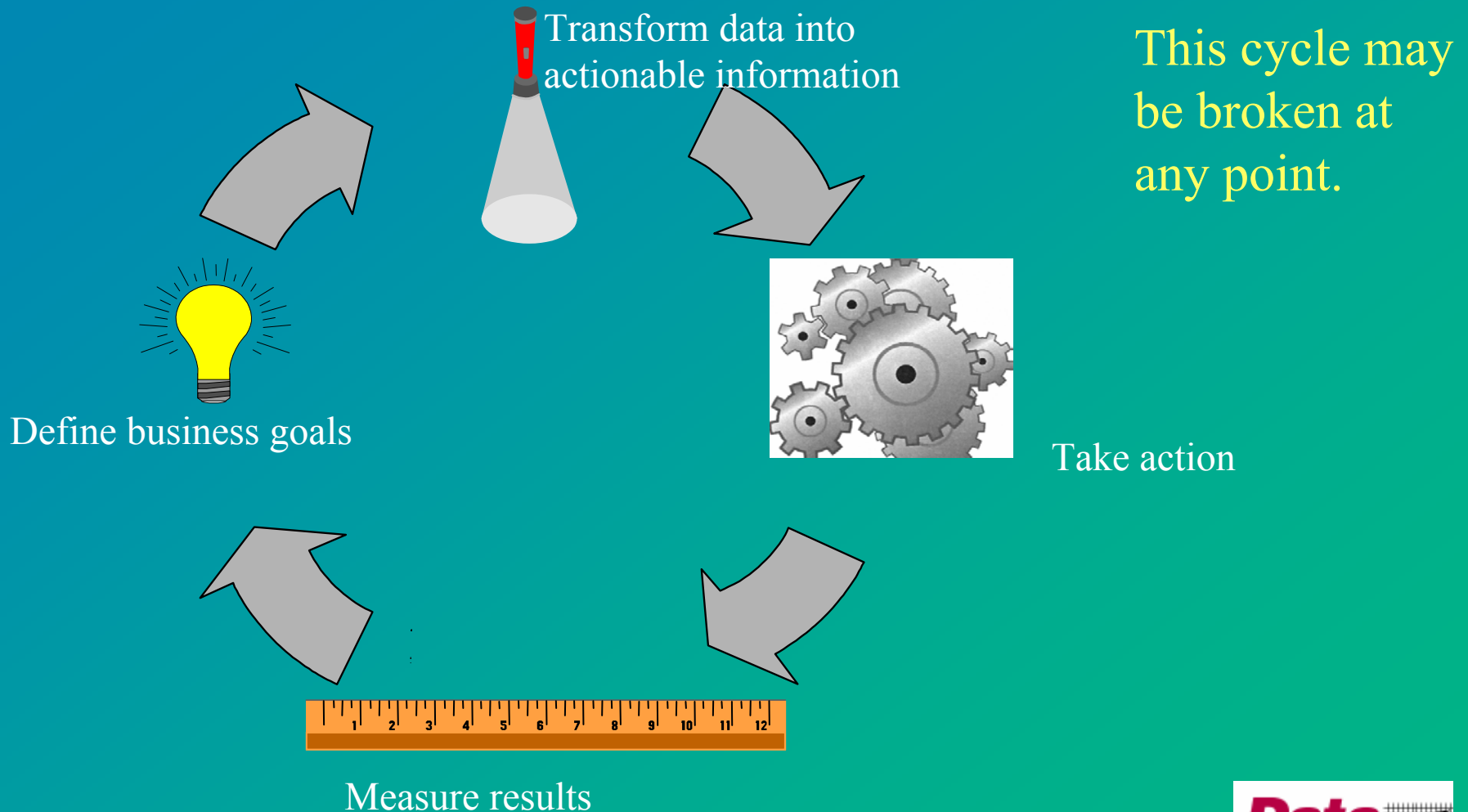
mjab@data-miners.com



Today's Talk

- ◆ Practical issues
 - What makes data mining difficult?
- ◆ Examples of Data Mining in Practice
 - Clustering towns for the Boston Globe
 - Analyzing prospective customer value for subscription-based businesses
 - Picking the best next offer

The Virtuous Cycle of Data Mining



Problems Defining the Business Goal

When has a customer left?

- When she cancels her card?
- When she hasn't made a charge in 3 months?
- When the balance does not support the cost of carrying the customer?



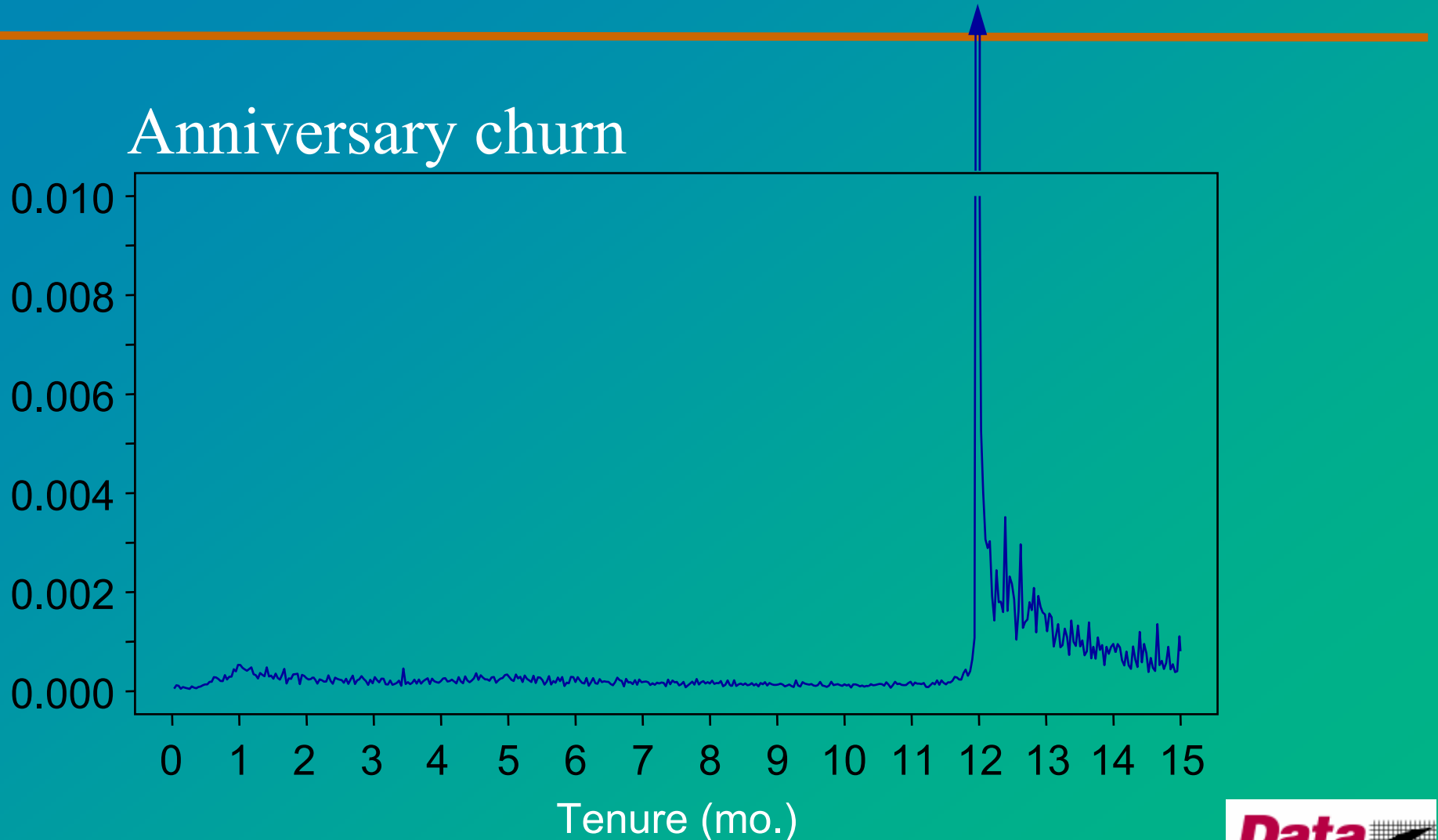
Problems Defining the Business Goal

Who is the yogurt lover?

- One answer for coupons
- Another for advertising



Problems Defining the Business Goal



Practical Problems with the Mining Itself

Learning things that aren't true

Learning things that are true, but not actionable

Interesting events are rare

Patterns in data reflect past management decisions

Technical issues can trap the unwary

Patterns May Not Represent Any Underlying Rule

- ◆ Some patterns reflect some underlying reality . . .
 - The party that holds the White House tends to lose seats in Congress at off-year elections.
- ◆ Others don't . . .
 - When the American League wins the World Series, Republicans take the White House.
 - When the Redskins win their last home game, the incumbent party keeps the White House.
- ◆ Sometimes, it is hard to tell without analysis . . .
 - In U.S. presidential contests, the taller man usually wins.

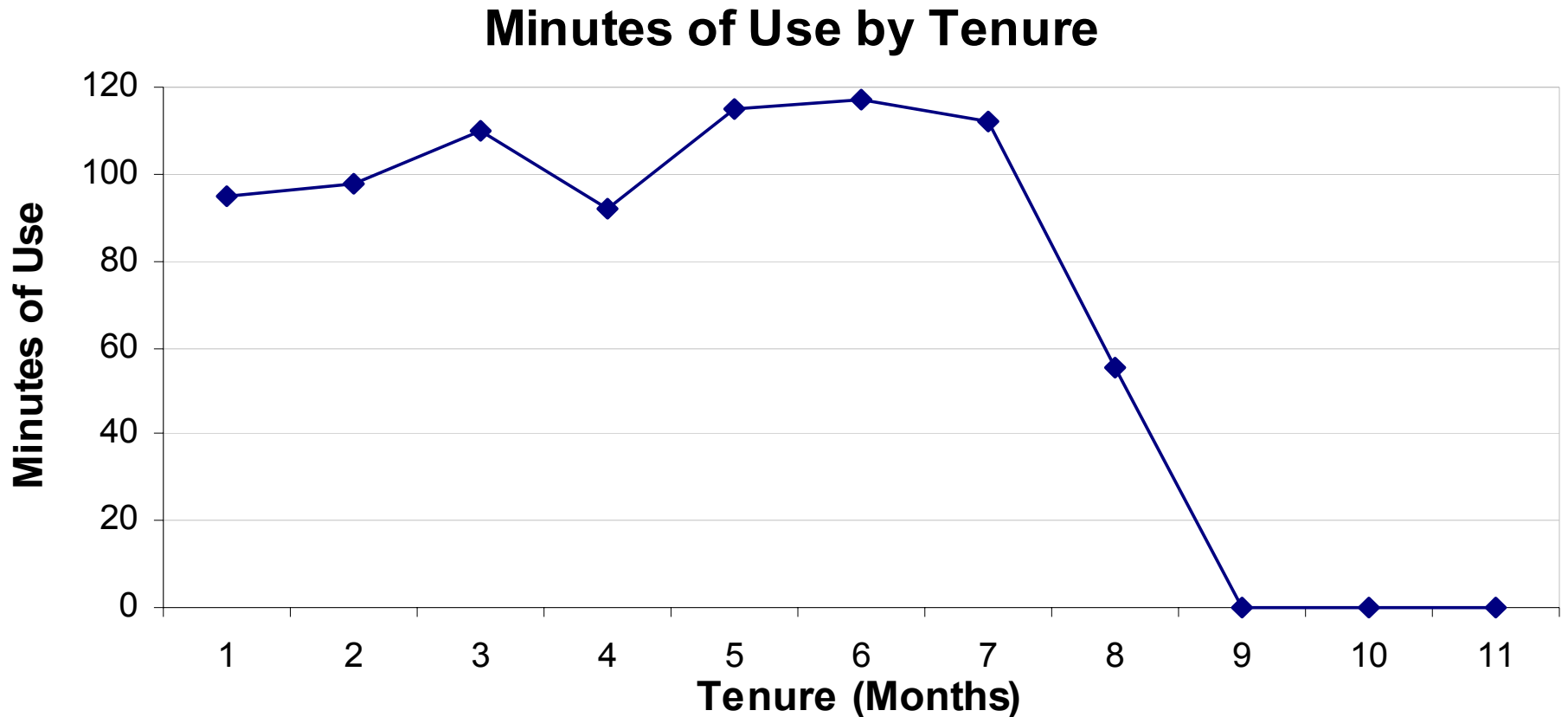
Correlation does not imply causation.

The Model Set May Not Reflect the Relevant Population

- Customers differ from prospects.
- Survey responders differ from non-responders.
- People who read e-mail differ from people who don't read e-mail.
- Customers who started three years ago may differ from customers who started three months ago.



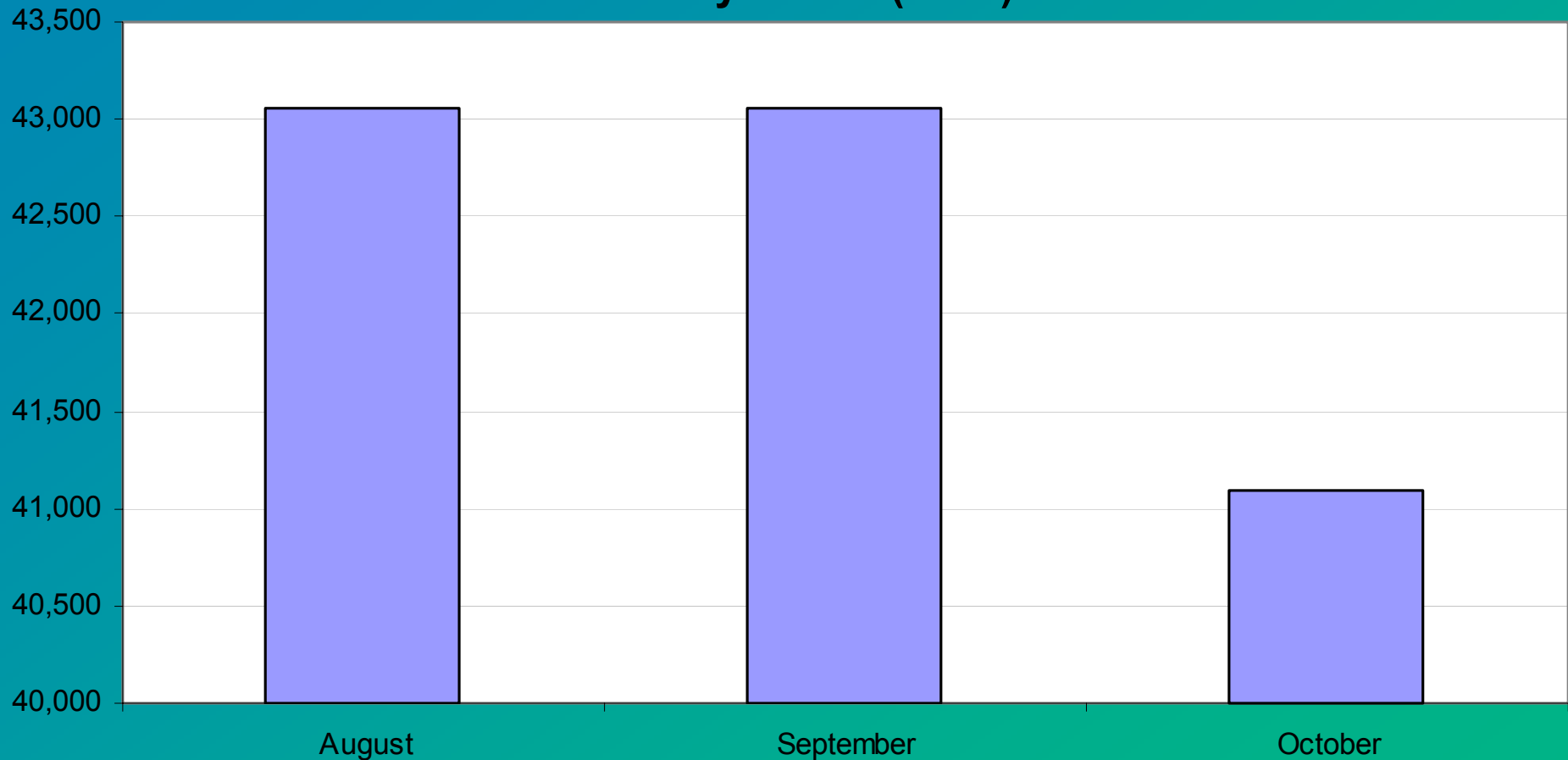
Learning Things That Are Not True



Does declining usage in month 8 predict attrition in month 9?

Learning Things That Are Not True

Sales by Month (2003)



Learning Things That Are Not Useful

- ◆ Results that arise from past marketing decisions
- ◆ Results that you already know
- ◆ Results that you already should know
- ◆ Results that you are not allowed to use

Patterns in Data Reflect Past Decisions

- ◆ People with voicemail make frequent calls to their voicemail number
- ◆ Everyone who has call forwarding also has call waiting

What would happen if a bank had a history of redlining?

Problems Acting on Results

Data mining project funded by group with no authorization to act on results

We built a model to predict who would leave
– The people predicted to leave, left in droves

A successful project?

Data Mining May Suggest Actions Counter to Corporate Culture

- ◆ Recommend what customer wants or what business wants to push?
- ◆ Recommend a less profitable rate plan?
- ◆ Change suspension policy?

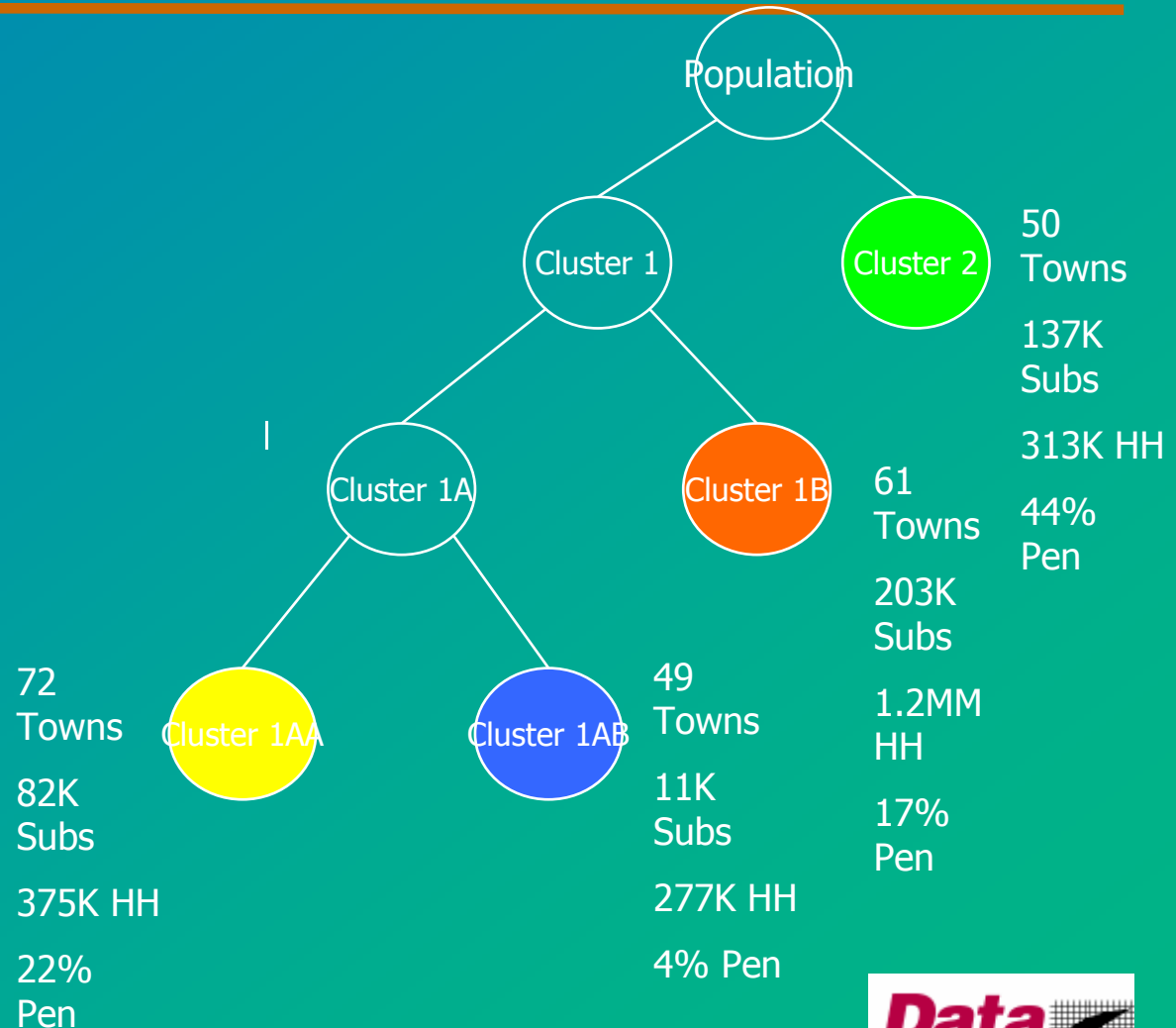
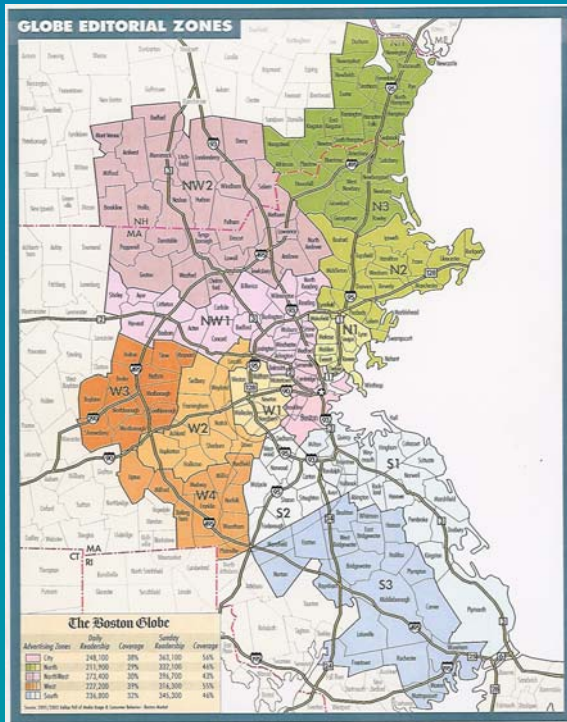
Data Mining in Practice

Clustering towns around Boston

The Business Goal

- ◆ The Boston Globe publishes regional editions with slightly different editorial content.
- ◆ There are constraints on the definitions of the regions.
 - They must be contiguous so trucks can drive a sensible route.
 - They must align with advertising regions.
- ◆ Within these constraints, make towns within a zone as similar as possible.

Clusters Based on Demographics



Clusters and Zones

Town Name	Editorial Sub-Zone	Home Delivery Subs	Penetration			Cluster
			Rate	Rank	Class	
Brookline	City	7,805	27.6%	81	1	2
Boston	City	37,273	14.6%	141	1	1-B
Cambridge	City	9,172	19.2%	112	1	1-B
Somerville	City	5,395	15.8%	132	1	1-B

3 out of 4 cities in the City zone are in cluster 1-B. Brookline is the exception.

Move Brookline to West 1?

Town Name	Editorial Sub-Zone	Home Delivery Subs	Penetration			Cluster
			Rate	Rank	Class	
Needham	West 1	4524	0.39%	44	2	2
Newton	West 1	18345	0.56%	9	2	2
Wellesley	West 1	2387	0.26%	87	1	2
Weston	West 1	2389	0.56%	7	2	2
Waltham	West 1	5894	0.25%	90	1	1-B
Watertown	West 1	4541	0.29%	76	1	1-B

Move Waltham and Watertown to City?

Case Study Conclusions

- ◆ The editorial zones designed by the editors already tended to be fairly homogeneous in terms of the demographic clusters we found.
- ◆ Clustering helped show what the “good” towns have in common.
- ◆ Clustering shows how the demographic purity of the editorial zones can be increased.

Clustering does not say whether that would be desirable.

Data Mining in Practice

Another Approach to Customer Value

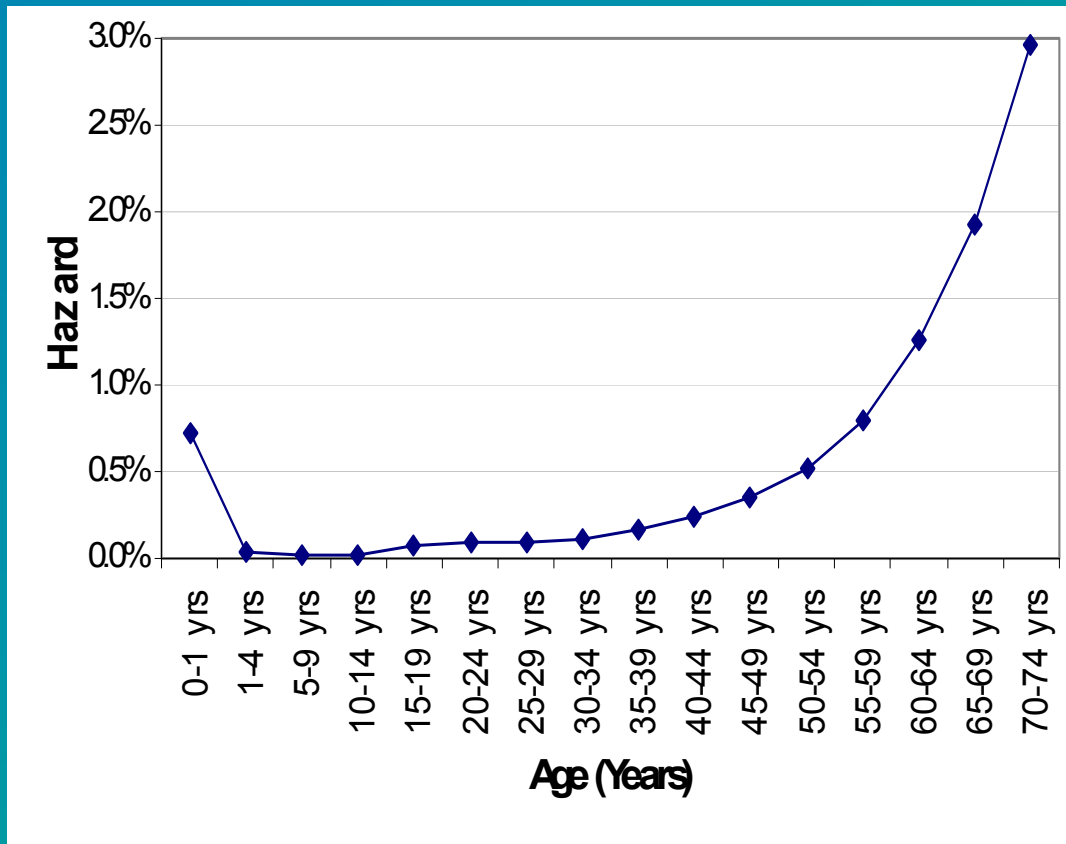


The Business Goal

- ◆ How much better are customers from a company store than customers from a dealer?
- ◆ How much better are gold card holders than green card holders?
- ◆ How much better are customers who signed up on the web than customers who responded to telemarketing?

Survival analysis can answer these questions.

Survival Analysis Begins with Hazards

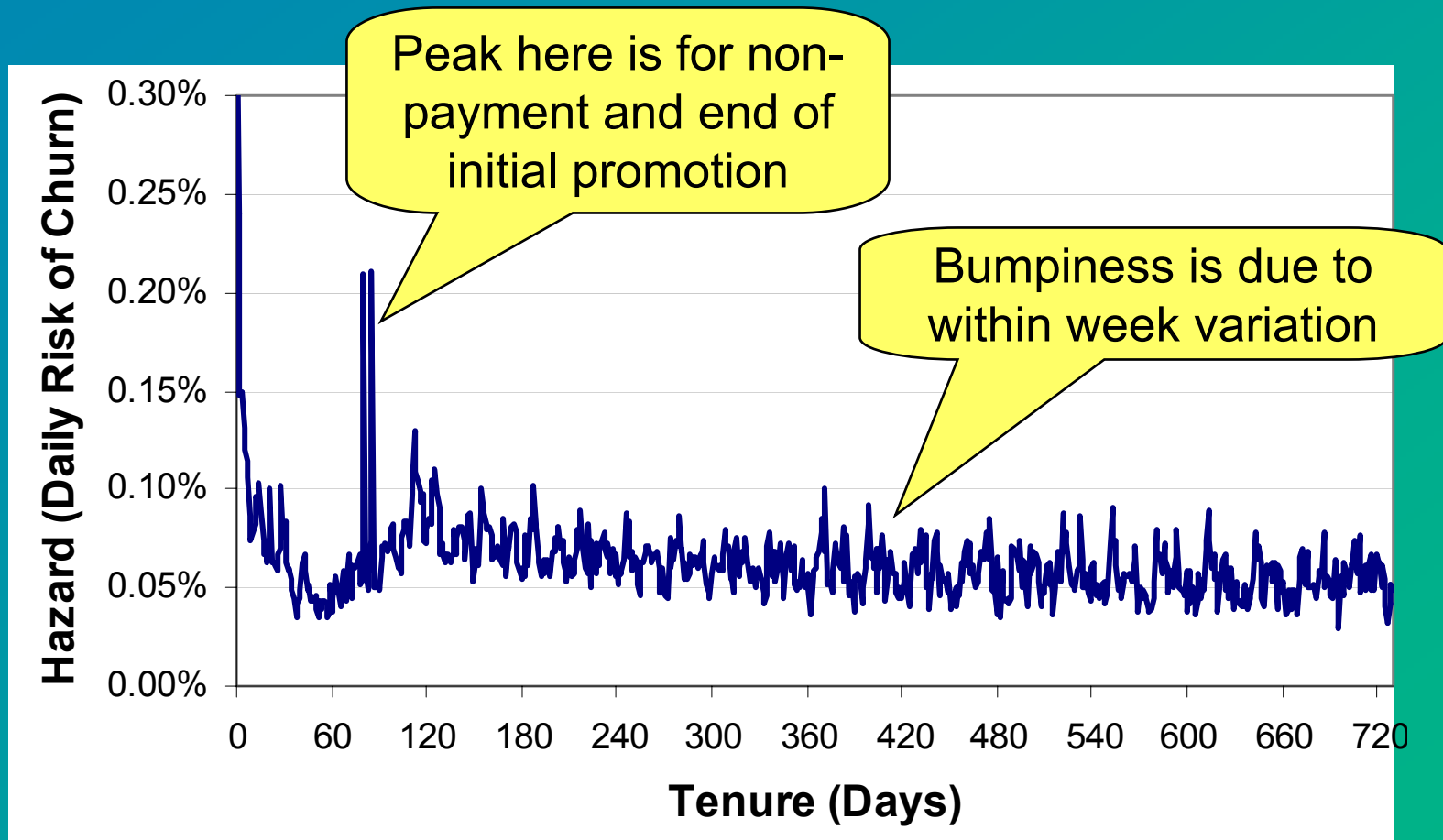


- ◆ Bathtub hazard starts high, goes down, and then increases again
- ◆ Example from US mortality tables shows probability of dying at a given age

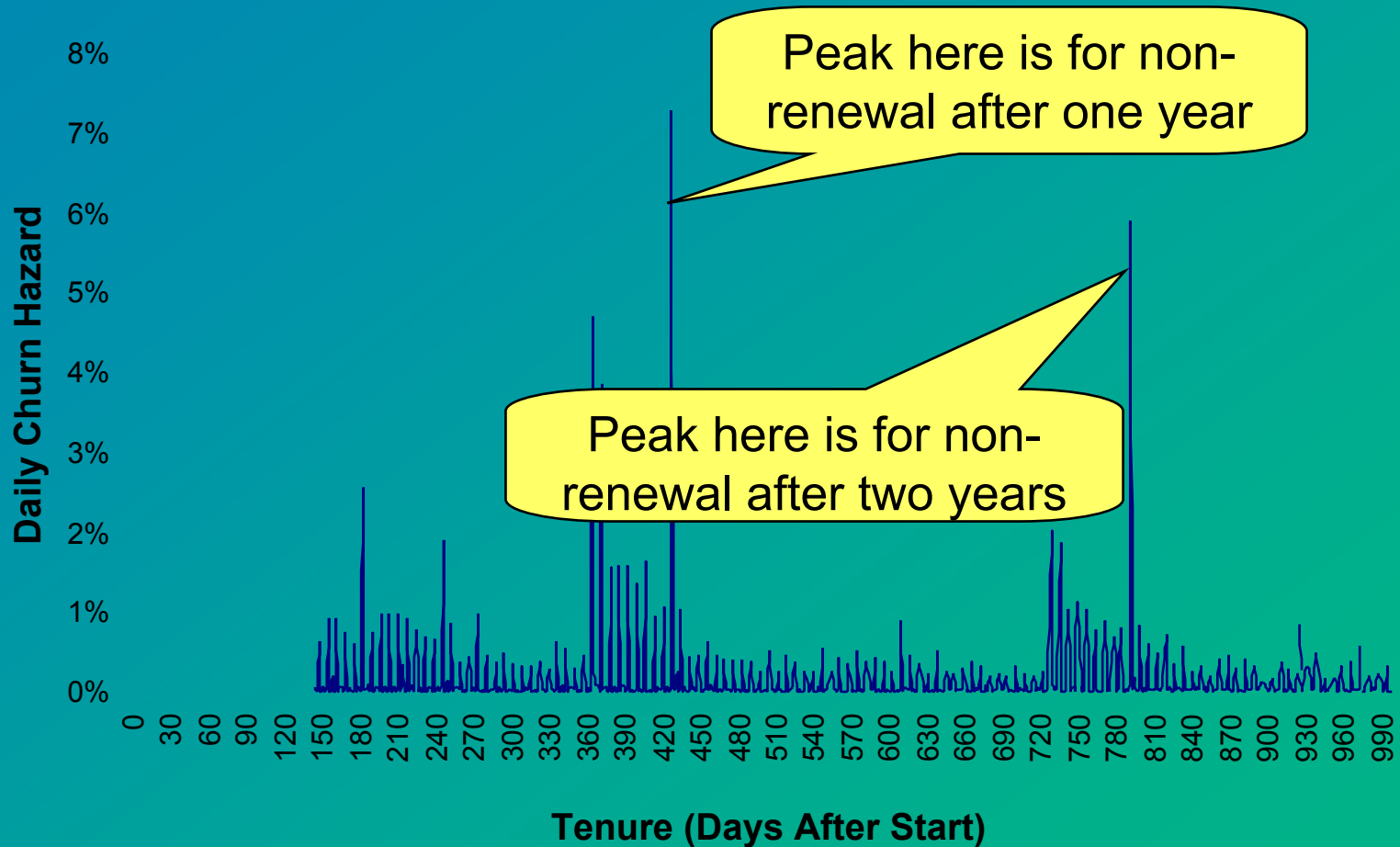
Calculating Hazards

- ◆ Hazard, $h(t)$, at time t is the probability that a customer who has survived to time t will not survive to time $t+1$.
- ◆
$$h(t) = \frac{\# \text{ customers who stop at exactly time } t}{\# \text{ customers at risk of stopping at time } t}$$
- ◆ Value of hazard depends on units of time – days, weeks, months, years.
- ◆ Differs from traditional definition because time is discrete – hazard probability not hazard rate.

Hazards Are Like An X-Ray Into Customers



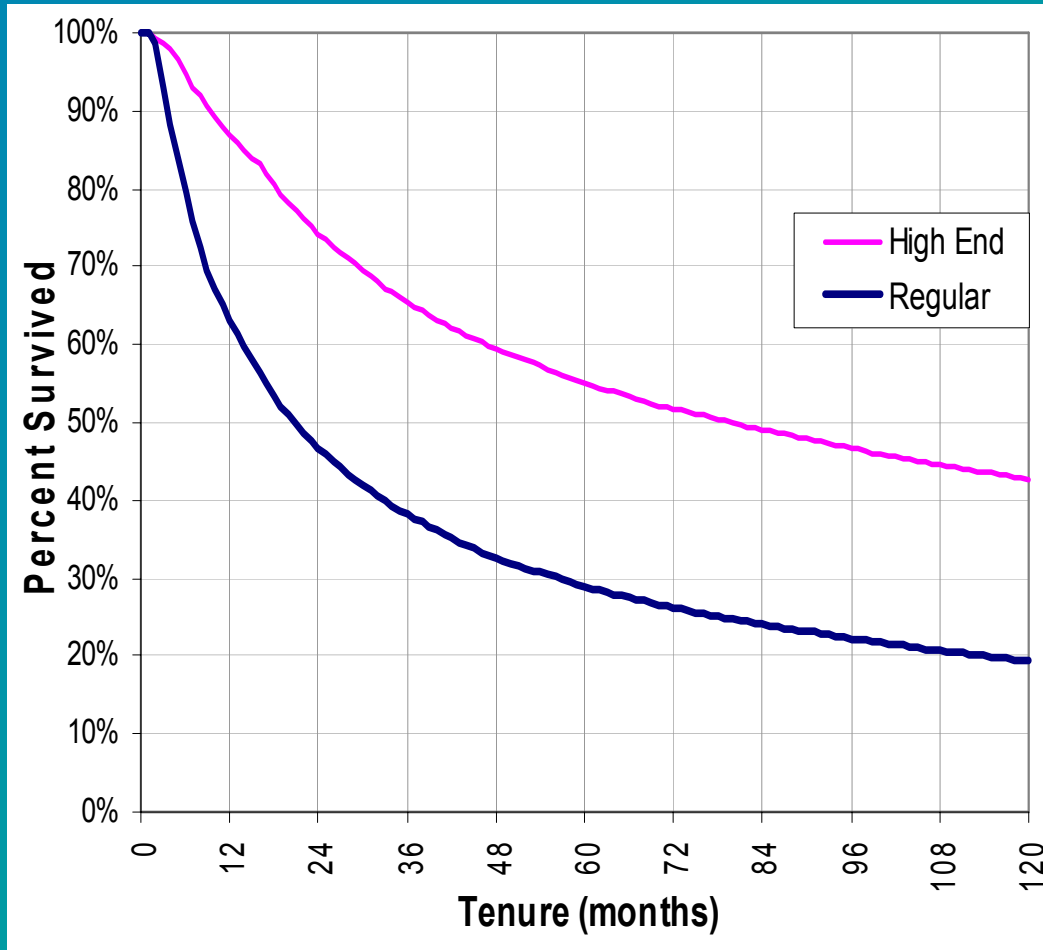
Hazards Can Show Interesting Features of the Customer Lifecycle



How Long Will A Customer Survive?

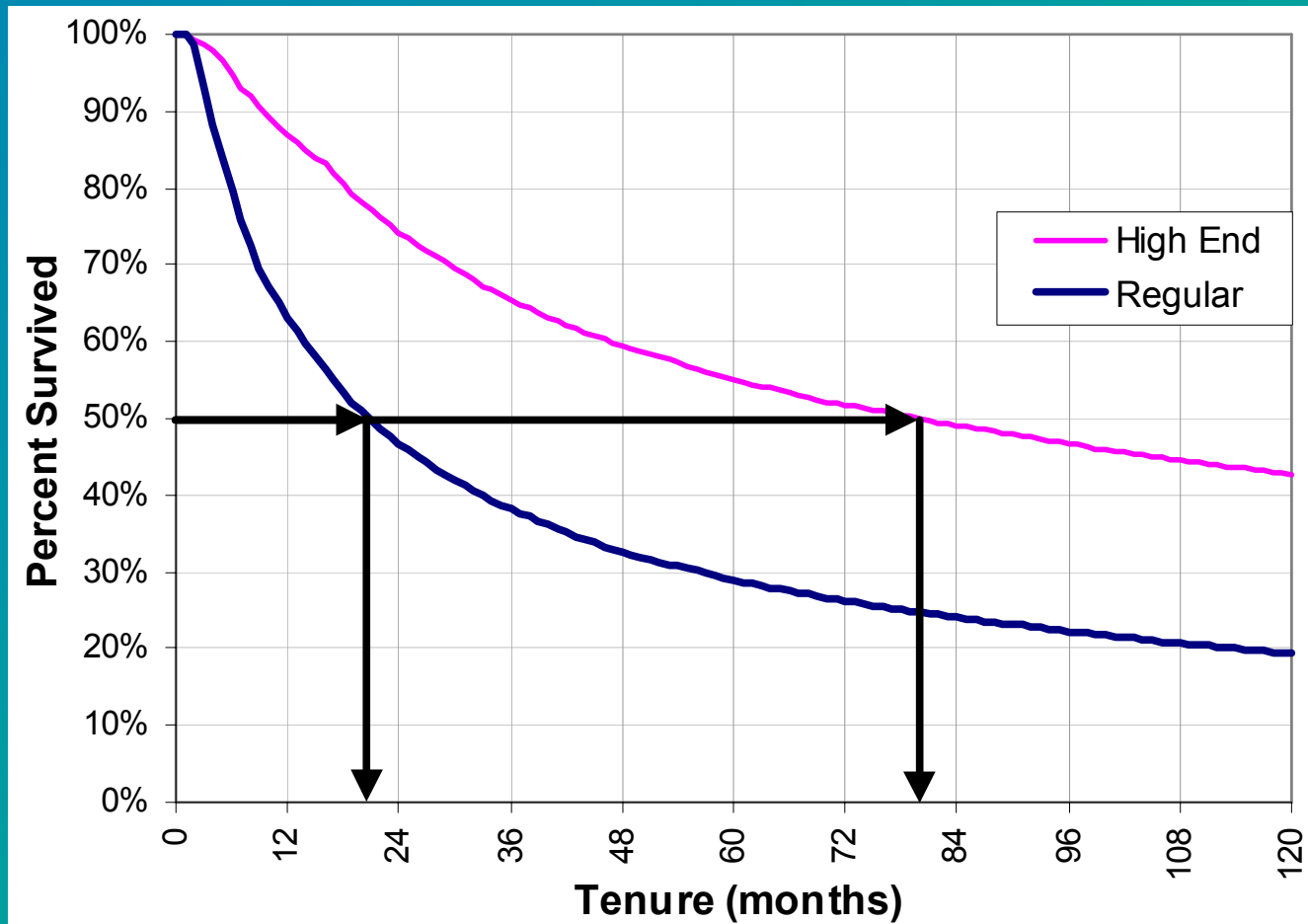
- ◆ Survival analysis answers the question by rephrasing it slightly:
 - What proportion of customers survive to time t ?
- ◆ Survival at time t , $S(t)$, is the probability that a customer will survive exactly to time t
- ◆ Calculation:
 - $S(t) = S(t - 1) * (1 - h(t - 1))$
 - $S(0) = 100\%$
- ◆ Given tenure data, hazards and survival can be easily calculated in SAS using PROC LIFETEST.

How Long Will A Customer Survive?

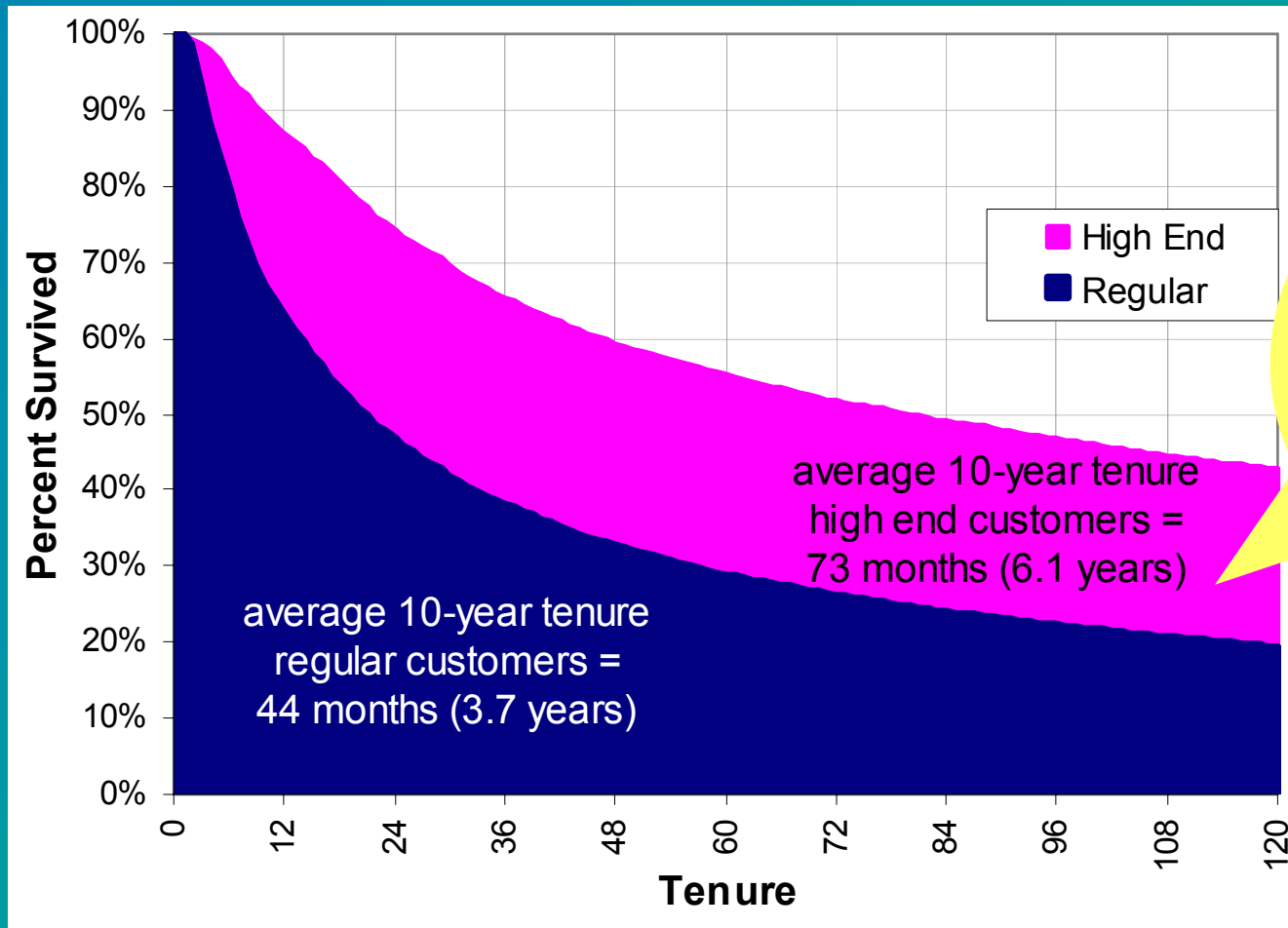


- ◆ Survival always starts at 100% and declines over time
- ◆ If everyone in the model set stopped, then survival goes to 0; otherwise it is always greater than 0
- ◆ Survival is useful for comparing different groups

Median Lifetime is the Tenure Where Survival = 50%



Average Tenure Is The Area Under The Curves



How much
is this
difference
worth?

Customer Value

- ◆ High-end customers pay \$50/month
- ◆ Low-end customers pay \$35/month
- ◆ High end customers last 29 months longer on average.

$$\$15 * 29 = \$435$$

Data Mining in Practice

Finding the Best Next Offer
for an On-line Bank

The Client

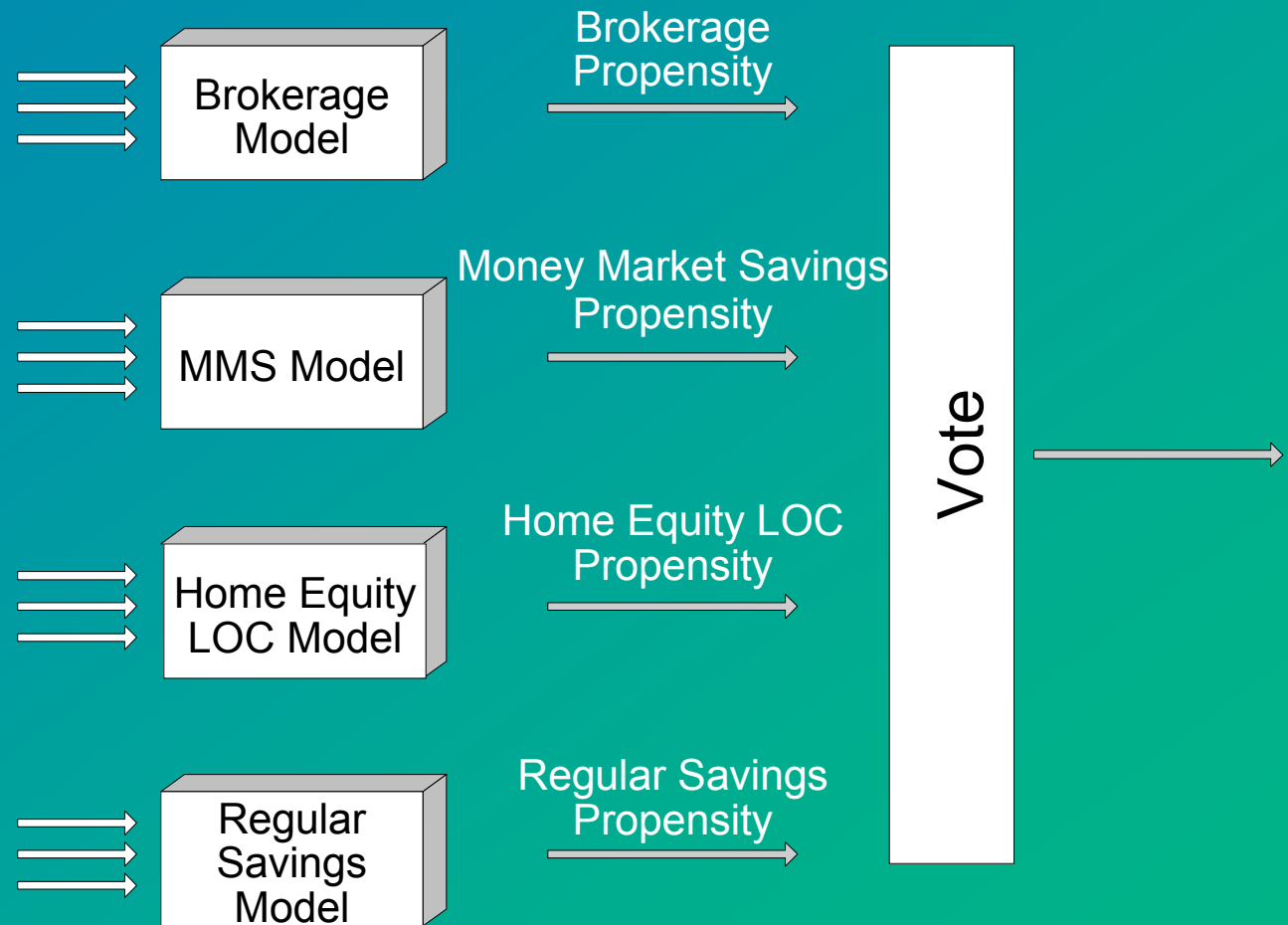


- ◆ One of the top 10 U.S. banks
- ◆ Web-based business
 - About 1.2 million accounts at time of study
 - Somewhat fewer customers because some have more than one account. (That is the point!)

Business Goals

- ◆ Increase profitability of on-line banking customers
 - by getting them to use more financial services.
- ◆ Increase loyalty of on-line banking customers
 - by being responsive to their needs
 - by encouraging them to think of the bank as their one-stop shop for financial services.

The Approach



If it walks like a duck . . .

"When I see a bird that walks like a duck and swims like a duck and quacks like a duck, I call that bird a duck."

--Richard C. Cushing

or

"Birds of a feather flock together."

--Anonymous

Problems With Profiling

- ◆ Current customers reflect past marketing decisions and data mining algorithms will find the patterns from past marketing campaigns
 - There is a risk of perpetuating past discrimination
- ◆ Owning the product changes customer behavior
 - People with CDs have low savings account balances.

The Data

◆ Household-level data

- Household income, ZIP code, number of people in household, total of all balances, months on line, presence of business account, ...

◆ Product-level data

- Presence or absence of each product with balances and tenure

◆ Transaction summaries

- Number of ATM transactions, number of credits, number of debits, ...



First Model: Brokerage Account

Train a model to recognize people with brokerage accounts.

- Customers who resemble brokerage customers, but who do not have a brokerage account, are good prospects.

Correlation of Brokerage with Other Variables

- ◆ The first step was to rank all variables by the extent of their correlation with the target variable
 - Being a Private Banking customer is the most important variable for predicting whether someone has a brokerage account
- ◆ **Problem:** Private Banking Customers do not get spam
 - All communication with Private Banking customers must be made through the private banker

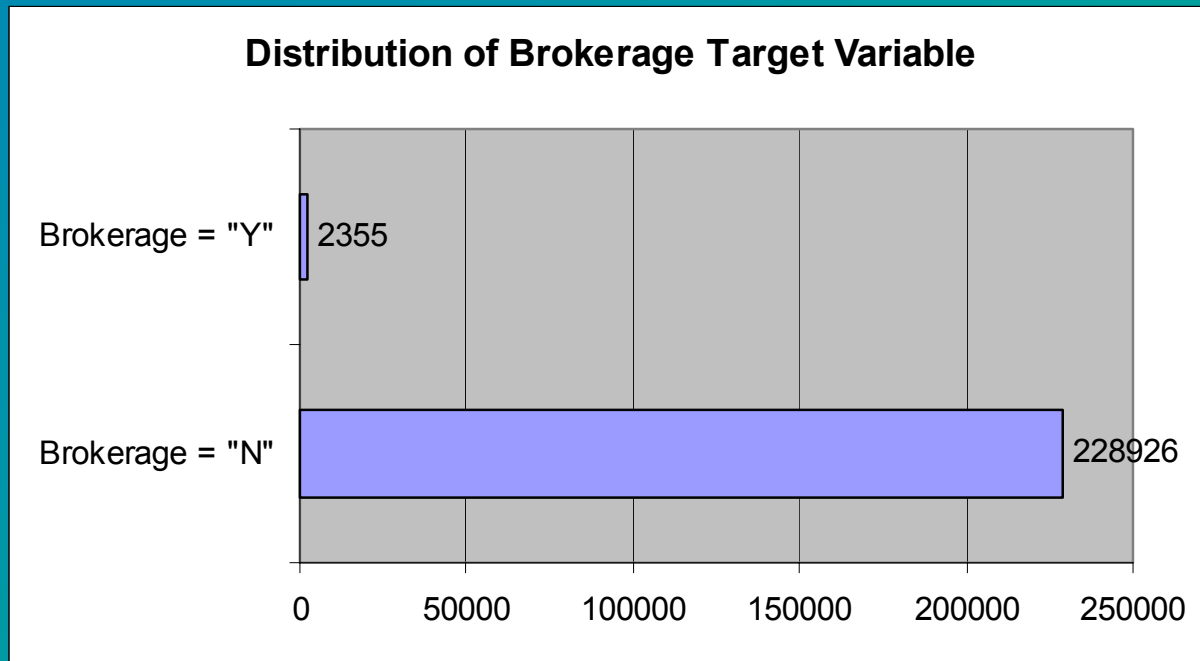
First Attempt to Build a Decision Tree (Only One Leaf)



As a classifier, this tree has an extremely low error rate: 1.2%.

As a customer segmentation tool, it is useless!

Too Few Examples



Adjust Scores for Oversampling Rate

- ◆ Different propensity models have different densities and will have different oversampling rates
- ◆ Need to map densities back to those of the population

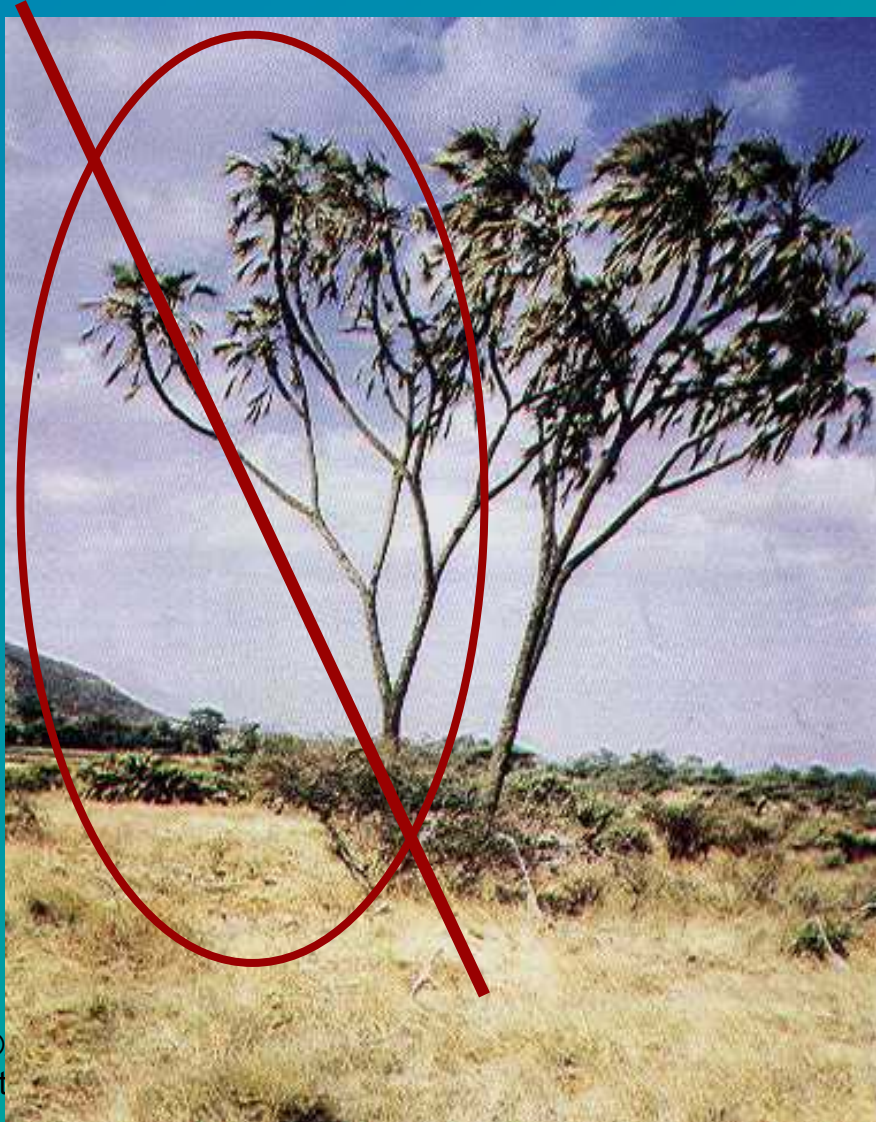
Cross-sell models depend on comparisons between models.

A Better Tree



As expected, the root node is Private Banking and the best leaves for brokerage are on the Private Banking side of the tree.

A More Useful Subtree



Chopping off the Private Banking side of the tree yields a model with decent lift and practical value.

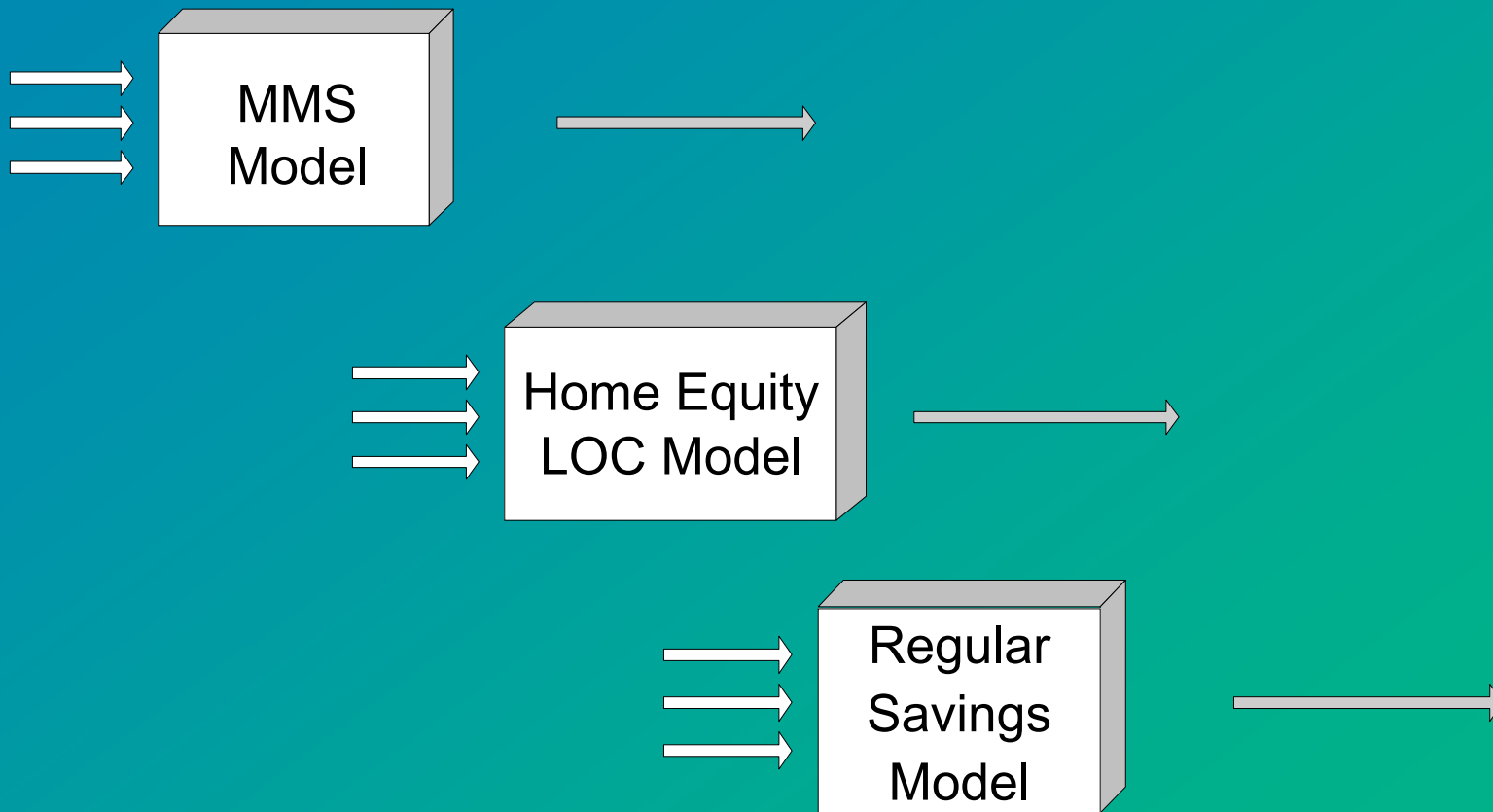
Measuring the Brokerage Model

- ◆ Three experimental groups of 10,000 each:
 - the modeled group
 - » modeled to respond and sent e-mail
 - the control group
 - » received same targeted e-mail as the modeled group
 - the hold-out group
 - » no e-mail.

Comparing Predicted to Actual Lift

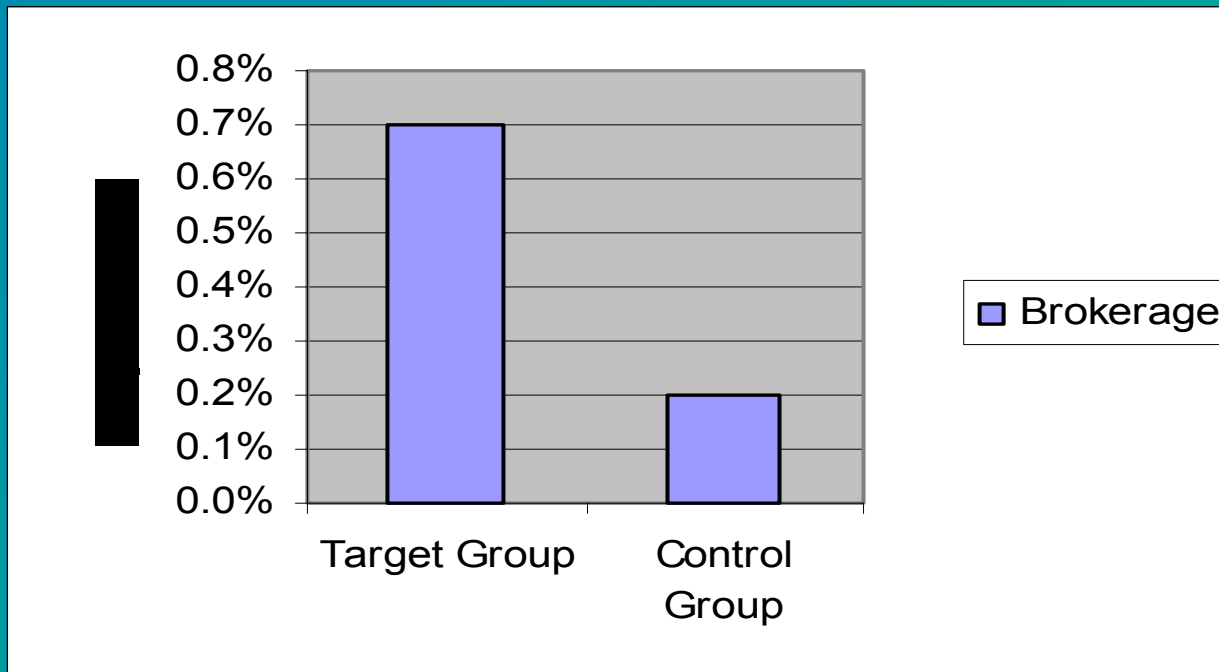
- ◆ Predicted lift: 2.3
- ◆ Model Group, Control Group, Hold-Out Group
 - 10,000 people each
- ◆ Model group response was 0.7%.
 - Actual lift of 3.5 albeit from a low base
- ◆ Control group response was 0.2%.
- ◆ Hold-out group response was 0.05%.

Repeat the Exercise for Three More Models

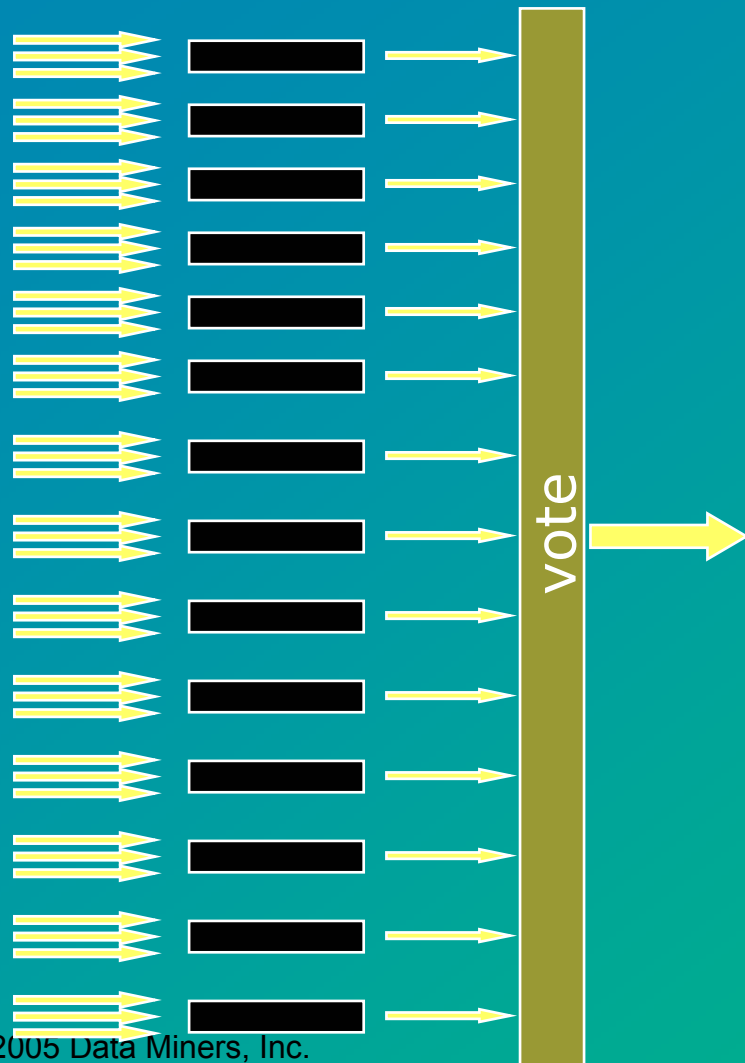


Results of Test Mailing

- ◆ Brokerage Q4 E-mail Campaign
 - Brokerage actual lift 3.5



Combining More than Thirty Models into a Best Next Offer Model



- ◆ Multiple models are built using the same input data but different weights and different derived variables.
- ◆ A score for each model is added to every customer signature.
- ◆ A customer's highest scoring product is her best next offer.

Further Reading

