Optimizing Mailing Lists and Content in Direct Marketing

Dr. Dan Putler, Alteryx
Agenda

• An introduction to Southern States Cooperative
• The business objective and decisions to be addressed
• The data used in the analysis
• The models developed
• Jointly optimizing the offer and the recipients
• A preliminary analytical app based on the analysis
• Questions
Southern States Cooperative

• Founded in 1923
• One of the largest farmer-owned cooperatives in the United States
• Owned by over 300,000 farmer members
• Purchases, manufacturers or processes feed, farm supplies, fuel
• 1200+ retail locations in 23 states
The Situation

Goal:
Better target, produce, and generate incremental profits from direct marketing initiatives.

Challenges:
1. Inability to bring together customer and marketing data from multiple sources.
2. No consistent tools to drive marketing analytics.
3. A lack of sophisticated analytics.

Solution:
Predictive analytics and execution optimization to improve the performance of campaigns and enable better business decisions.
A Need to Simplify, Quantify and Analyze

“Overall, we were looking for a way to simplify the process of extracting the data, and then turn that data into a usable form for our marketing analysis.”

Greg Bucko
Manager of Customer Insights
Southern States Cooperative
The Business Objective and Decisions to Address

- Southern States Cooperative distributes a number of direct mail seasonal catalogs to different segments each year.
- Southern States Cooperative was selecting customers to include in the mailing list for a catalog using a set of “rules of thumb”.
- Needed to maximize the incremental gross margin from the catalog net of mailing costs for the Fall catalog.
- Business decisions to be addressed:
  - Which of the Southern States Cooperative current customers should be sent the fall livestock catalog?
  - What Stock Keeping Units (SKUs) should be included in the catalog?
Solution Overview

1. A probability model of catalog use
2. A revenue model to assess the incremental gross margin that positive customer catalog use provides
3. An estimate of expected gross margin percentage for each customer
4. An optimization module to assist direct marketing managers in selecting the mailing list and catalog items
Data and Predictive Models
The Available Data for Model Development

- Southern States Cooperative historical customer transaction records
- The list of SKUs (items) included in both the Fall 2010 and 2011 livestock catalog
- Third-party (Farm Market iD) farm level “firmographics”
- Geocoded customer locations
- Geocoded locations for both Southern States Cooperative and Tractor Supply outlets
The 2010 Livestock Catalog Calendar

July 5
Creation of customer mailing list for fall livestock catalog begins

Aug 15
Catalog arrives to approximately 35,000 customers

Sept 30
The last day catalog pricing is guaranteed
The Measures Used

- **RFM**
  - **Recency**: The number of days prior to July 5, 2010 that a customer made a purchase at a Southern States outlet.
  - **Frequency**: The number of different days a customer purchased at least one item from a Southern States outlet between July 5, 2009 and July 4, 2010.
  - **Monetary Value**: The dollar value of all purchases made by a customer between July 5, 2009 and July 4, 2010.
  - The number of unique catalog SKU / date combinations in a customer’s purchase history between July 5, 2009 and July 4, 2010.
  - Customer level gross margin percentage based on a customer’s purchase history.
On the Cutting Room Floor

• Farm/ranch characteristics
• The drive time to the nearest Southern States and Tractor Supply outlets
Results of the Probability Models

• Logistic regression, decision tree, and random forest models were developed. Logistic regression is the preferred model.

• The RFM measures and the measure of past purchases of catalog items are all statistically significant in the probability model.

• The most important factor in driving customer use of the catalog is past purchases of items contained in the catalog.

• All factors have effects that are of the “diminishing returns” variety.

• The overall model fit is better than is typically the case for these types of models.
The Incremental Revenue Model

• The model is highly structured to allow for a clear estimate of the effect of customer catalog usage on revenues from that customer.

• The model is estimated using four different periods to establish both expected baseline revenues and the level of incremental revenue that can be attributed to a customer responding to the catalog.

• RFM measures, time period and catalog usage indicators are used as the predictors

• Linear regression is the modeling algorithm used

• The model fits extremely well, and indicates that customers who buy catalog items in the catalog period spend 30% more than they would otherwise.
Direct Mail Optimization
The Optimization Problem

\[
\arg\max_{s \in S, c|s \in C} \sum_{i \in c|s} \Pr(s, s_i, d_i) IC(d_i) - M
\]

subject to: \( ||s|| \leq N \)

\( S \) is the set of all possible catalog items
\( s \) is a subset of catalog items that “fit” into the catalog
\( N \) is the maximum number of items allowed in the catalog
\( C \) is the set of potential customers who could receive the catalog
\( c|s \) is the subset of customers that have an expected positive return, conditional on the set of items in \( s \)
\( \Pr(s, s_i, d_i) \) is the probability that customer \( i \) will respond to the catalog which depends on \( s, s_i \) (past items purchased), and \( d_i \) (other customer characteristics)
\( IC(d_i) \) is the expected contribution dollars associated with a positive response
\( M \) mailing cost
Observations about the Optimization Problem

- The solution for the optimal set of customers to mail to (given the items in the catalog) can be found using a break even condition.
- The selection of the items to be included in the catalog is a variant of the “knapsack” problem.
- The objective function is not well behaved, thus traditional numeric optimization methods (based on derivatives) cannot be used.
  - A genetic algorithm approach is used instead.
- Instead of solving the constrained optimization problem, a penalty function is added to the objective that embeds the constraint.
The Bottom Line: The 2010 Catalog and Model

Mailings:
- Actual: 34,546
- Model: 12,824 (a 63% reduction)

Positive responses:
- Actual: 6,908 (20% response rate)
- Model: 4,324 (34% response rate)

Estimated incremental gross margin less mailing cost:
- Actual: $157,366
- Model: $193,604 (24% increase)
The Bottom Line: The 2011 Catalog and 2010 Model

Mailings:
• Actual: 9,285
• Model: 4,733 (a 49% reduction)

Positive responses:
• Actual: 2,537 (28% response rate)
• Model: 1,846 (39% response rate)

Estimated incremental gross margin less mailing cost:
• Actual: $71,182
• Model: $77,863 (9% increase)

The number of model indicated mailings if the 2010 and 2011 mailing lists were combined:
• 18,951
• With catalog optimization: $303,164 incremental gross margin
Results Summary

1. With Alteryx modeling tools, Southern States Cooperative can improve the performance of its direct mail activity.
   - Model show that is should be able to increase gross margins minus mailing costs by 24%.

2. Smarter decision-making has sped up dramatically throughout the organization; marketing, finance, retail.

3. Data analysts have been freed to concentrate on higher-level, business process analytical tasks.

4. Southern States Cooperative has better insight into placement and competitive environment of its retail stores.
Next Steps

• Look at improving the measure of past catalog purchases
• Add a metric of for the value of each SKU included in the catalog
• Create an analytic app based on the module
• Conduct A/B testing to see if the approach really does live up to its implied promise
Q&A