Determine the Effectiveness of a Multi-channel Advertising Campaign

<u>Team 4</u>
Miguel Dieguez
Jack Sun Hom
Rahul Prabhu
Landric Walden
Anne Zeeman

Executive Summary

The advertising campaign was effective for generating new net sales. Overall, there was a 13.5% conversion to sale with a 14.1% advertising spend to sales ratio (Exhibit 8). These rates are considered successful in the marketing profession where a mailing response rate of 1% is the norm.

We segmented customers into three clusters: 1. "Valued Customers," 2. "Infrequent Customers," 3. "Occasional Customers." Valued Customers were responsible for approximately 58% of total sales, whereas 3% was from Infrequent Customers, and 36% from Occasional Customers. It is tempting to draw the conclusion that advertising to Infrequent Customers should be terminated in exchange for perhaps increased advertising to more responsive clusters, but this is ill advised. Instead marketing managers should gather more information, determining characteristics such as geographic location, demographics, and psychographics if possible. This information should be used to change the message, or marketing channels used with the clusters in order to increase conversion rates. Innovative marketing strategies, such as frequent buyer discounts, can be employed to induce these customers to buy. In addition, data collected on Valued Customers can be used to increase amount spent, as well as frequency.

Customers had the potential to receive four "touches" or pieces of communication from the retailer: catalog, postcard and two emails. The print campaign (catalog and postcard) targeted Valued Customers. Catalogs and postcards were the most effective touches of the campaign. Of the customers who received catalogs, 14.1% made a purchase, and for those who received the postcard, 17.6% made a purchase. Print advertising was definitely effective in generating sales. However, it must be noted that this part of the campaign was already skewed towards the customers who were most likely to buy: Valued Customers. Since this segment responded best to the print advertising campaign, we recommend that the company send catalogs, postcards and emails to all customers in this segment. Infrequent Customers seem to be neglected in the print advertising campaign, probably due to ROI concerns. Again, it is important that the Infrequent Customers cluster is looked at as an opportunity for growth, and not simply a less profitable segment.

The email campaign was split evenly across the clusters and not aimed at any segment in particular. Although the email campaign was not statistically significant, due to the low marginal cost, this "touch" of the campaign should be continued since it is mildly effective. In addition, it is difficult to say what the individual impact of an email is to consumers who received it. One of the major challenges to marketers today is campaign attribution. Customers receive multi-channel marketing messages and may choose to act on any of them. For instance, catalog and postcard distribution probably boosted Internet sales. USPS.com says that 37% of catalog recipients account for a retailers web site sales. Perhaps one touch has the singular impact needed to influence a sale, or maybe it's the nature of a series of touches that influences sales. Managers should track and control this data comparing it in time series to understand incremental increases in sales due to various touches of campaigns.

Technical Summary

Data exploration began by using bar charts to explore the composition of the dataset. Past consumer behavior – defined as frequency, recency, and monetary – was plotted and colored by new net sales (Exhibit 1). While the consumer data was relatively evenly distributed along the monetary dimension, there was a larger proportion of consumers with more recent purchases and either very high (i.e. 8x+) or very low frequency (i.e. 1x). These visualizations also revealed that new sales were more common among consumers with more recent purchases and a high frequency of purchases.

Similarly, bar charts were used to explore dataset composition with respect to the penetration of the advertising campaign. Two significant discoveries were made. First, there was a notable difference in penetration of the print campaign as compared to the email campaign. Of the 10,000 consumers in the dataset, 74% received catalogs and 55% received postcards while exactly 50% received the first email and only 40% received the second email (Exhibit 2). Second, and more importantly, a pattern was discovered in exploring which costumers received each element of the advertising campaign (Exhibit 3). Customers with more recent purchases and a higher frequency of purchases were much more likely to receive the catalog and postcard than customers with low scores (Exhibit 4).

It was through these data visualizations that customer segmentation started to become apparent. Cluster analysis was used to verify our intuition about which combination of consumer behaviors defined each customer segment and the relative number of segments. To effectively use cluster analysis within XLMiner, the data set was reduced to 4,000 records vice the original 10,000. This smaller data set was also used for classification trees, logistic regression and discriminant analysis.

During initial cluster analysis, we directed the model to produce four clusters based on recency, frequency, monetary and whether customers received the catalog, postcard, email #1 and email #2. Note that variables such as net sales, type of net sales, etc. were left out of the cluster analysis. The clusters are strictly based on previous customer behavior. After creating a few models, it became clear that in every four cluster model two clusters were always very similar in size and attributes so the number was reduced to three. The hierarchical model with three clusters resulted in customer segments that are sufficiently separated and easier to understand.

Cluster #1 is the largest cluster, containing 2480 records. It is made up entirely of customers who have purchased within the last 12 months. Their distribution of purchases within the 12 months decreases as the number of months increases. The largest group in the segment are those who purchased in the last three months (recency of 7) and the smallest group were customers who purchased 9-12 months ago (recency of 4). Customers in Cluster #1 also were the most frequent purchasers. Customers who purchased 8 or more times and between 5 and 7 times were the most highly represented.

Cluster #2 was the smallest cluster, containing only 648 records. This customer segment was predominantly customers with recency scores less than 3, meaning their last purchase was more than 13 months ago. Frequency of purchases in Cluster #2 was the opposite of Cluster #1, with the majority having only purchased between 1 and 4 times. Cluster #3 is only slightly larger than Cluster #2, containing 871 records. Of these, recency scores were all above 4, meaning purchases within the last 12 months but also had the opposite distribution as seen in Cluster #1. Most customers in this segment purchased between 6-12 months ago versus in Custer #1 where most customers purchased between 0-6 months ago. The frequency of purchases in Cluster #3 is predominantly in the 1 to 4 times range with zero records for 8 or more purchases (frequency of 4).

Given the explanatory goal of this project, classification trees, logistic regression, and discriminant analysis were considered in developing a model for this dataset. Initial models, however, yielded poor results since they were unable to accurately predict any "Yes" cases. The reason was that the success class – defined as "Yes" for new net sales – comprised the minority in the dataset. More specifically, less than 14% of the 10,000 customers in the dataset made another purchase. In order to model this data, over sampling was performed to achieve a naïve rule of approximately 33%, or twice as many "No" records as "Yes" records.

Classification Tree

Initial classification trees always predicted "No" sales for all models tried even with the over sampled data because for every value of the predictor, the chance of "No" sales was higher than "Yes" sales. In order to gain insights into the data, the size of the data had to be further reduced until the ratio between "Yes" Sales and "No" Sales was 40% / 60%.

Several models of the classification tree were run and the best fit was found in the model ran with all behavioral attributes and advertising campaign methods included (Exhibit 5). The results were in-line with the findings from data exploration. Observing the model: Customers who received postcards, bought recently (in the last 8-9 months), received the email#1 and buy frequently were most likely to buy. As stated earlier, the print campaign was focused towards customers who bought recently and bought frequently. The classification tree confirms the advertising campaign was well directed to the customers who were most likely to buy (Cluster #1) and sending the first email improved the odds of getting a customer to buy even further. Also, the monetary behavioral attribute remains insignificant and sending a postcard seems to improve the chance of the customer buying over just sending a catalog (Note – as per the data, most customers who received a catalog, did receive the postcard).

Logistic Regression

Prior to oversampling the dataset, logistic regression models offered very little explanatory insight. Every model that was attempted was unable to predict, whether correctly or incorrectly, any "Yes" records. Consequently, the overall error rate for each model was only 13.5%. This was due to the error rate for the "No" class being zero while the error rate for the minority "Yes" class was 100%. In other words, the overall error rate exactly matched the proportion of the success class, or buyers, in the dataset.

After the data was oversampled, the explanatory effectiveness of the models improved. Initial analysis focused on whether new sales could be better explained by just past consumer behavior, just advertising campaign "touches", or all of those predictors combined. In addition, the effectiveness of using predictor scoring instead of categorical dummies was investigated. While some models yielded lower overall error rates, extra consideration was given to models that accurately predicted a greater number of "Yes" records. This is especially relevant since part of the goal of this project was to explain the customer segmentation that best responded to the advertising campaign.

The best logistic regression model (Exhibit 6) had an overall error rate of only 33.98%, but accurately predicted only 280 out of 1,355 buyers for an error rate of 79.37%. The model included the following predictors:

Recency_07-09 Mo.
Recency_10-12 Mo.
Recency_13-18 Mo.
Recency_19-24 Mo.
Monetary_$200+
Receive Catalog? (Y/N)_Y
Receive Postcard? (Y/N)_Y

Each of these predictors was significant at a 5% level of significance. The regression output indicates that whether a customer received the catalog had the largest odds factor followed by whether a customer received the postcard. The output also revealed that customers who purchased longer than six months ago were less likely to make a new purchase, and customers whose last purchase was greater than $200 were more likely to purchase again. These conclusions were supported by the team's collective domain knowledge and the personal experience of several team members. It is also important to note the absence of the recency predictor term for 4-6 months as well as all of the frequency predictor terms. Intuitively it would seem these should be included. However, these terms are autocorrelated with the predictor term for catalog received, since more than 90% of the customer segment defined by those predictor terms was sent the catalog.

Discriminant Analysis

Initial discriminant analysis seemed to indicate the best ability to identify buyers compared to other methodologies. However, in order to verify our thoughts, oversampling was implemented by eliminating almost 60 percent of the non-buyer data. Many models using discriminant analysis were run and the smaller data set provided slightly more conclusive results, with all three advertising campaigns having a positive impact to varying degrees. Depending on the variation of predictors used, error rates when trying to correctly identify buyers ranged from 28% to 35%, and error rates when trying to properly identify non-buyers ranged from 44% to 52%. The best overall error rate was approximately 39% (Exhibit 7). We assume that the cost of misclassifying a buyer as a non-buyer was significantly more damaging for a business versus the opposite, due to the fact that there is a lost business opportunity (versus the cost of advertising expense to someone who does not buy). When increasing the cost, the error rates increased significantly.

When performing discriminant analysis, the largest concern when comparing relative strength of predictors is the difference in the predictor's value between success and non-success classes. This difference ultimately affects the analysis' outcome and classification of a customer as a buyer (success) or non-buyer (non-success). In our model's case, the most significant predictors in order of relevance were if a customer received a catalog, if a customer received a postcard, and if a customer's last purchase was more than $200 (Exhibit 7). Controlling for past purchasing behavior, emails also increased the likelihood of a customer making a new purchase. However, the effects were significantly less. The difference between the coefficients of the success and failure classes was very small. Although the model's overall error rate (39.7%) was high, and the error was better than logistic regression and more accurately predicts that success class. Those who bought more recently and more frequently, received a catalog, and received a postcard were most likely to accurately be classified as a buyer.

When determining the effectiveness of each campaign, the company must consider two separate but related cost issues. First, the actual cost of each advertising approach differs. The cost of developing, printing, and shipping a catalog to a customer is more expensive than the cost of sending an email. Second, the cost of misclassifying customers as non-buyers is higher than misclassifying them as buyers. While it was not possible to isolate the individual effectiveness of each advertising channel, these costs should be evaluated relative to the overall effectiveness of the campaign (Exhibit 8).

Exhibit 1 – Past Consumer Behavior

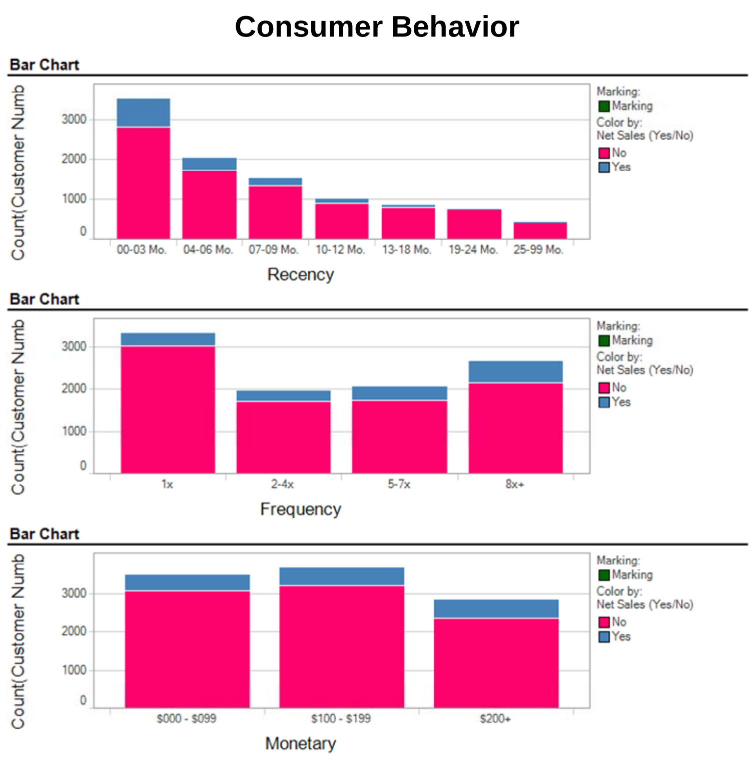


Exhibit 2 – Advertising Campaign “Touches”

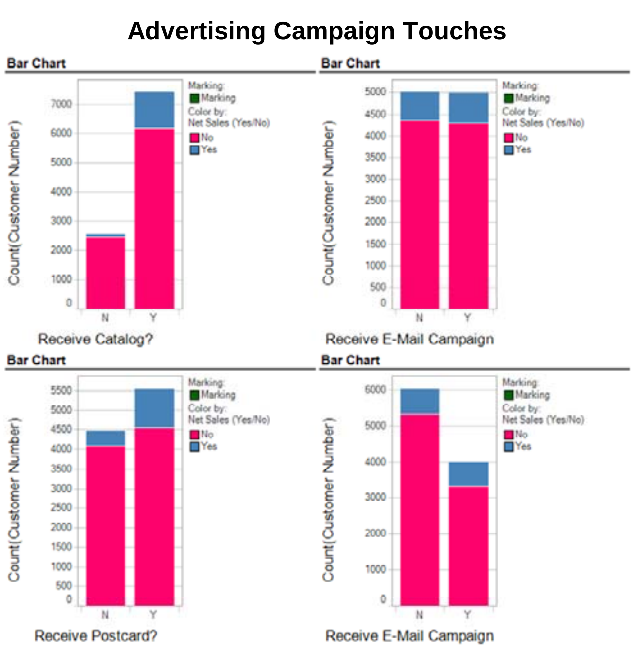
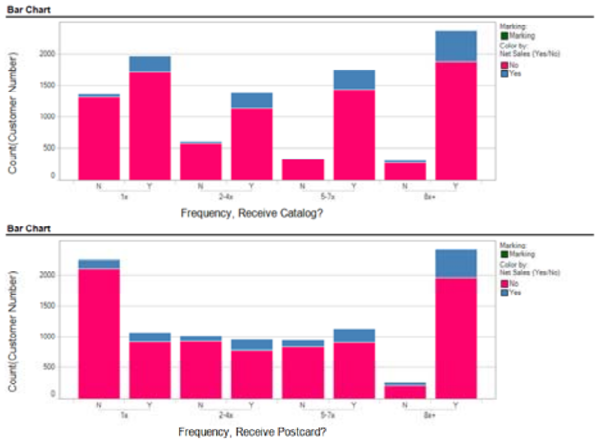
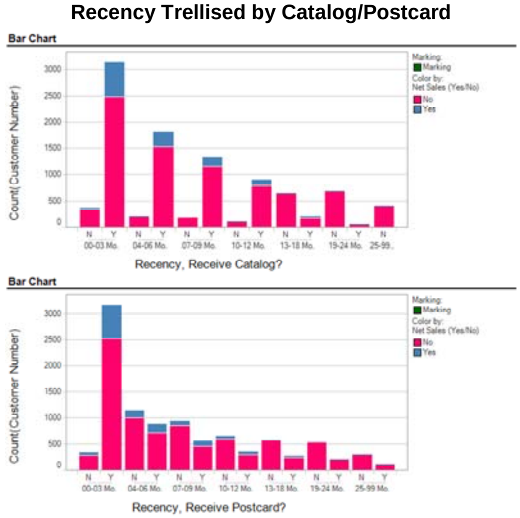


Exhibit 3

Frequency Trellised by Catalog/Postcard



Recency Trellised by Catalog/Postcard



Monetary Trellised by Catalog/Postcard

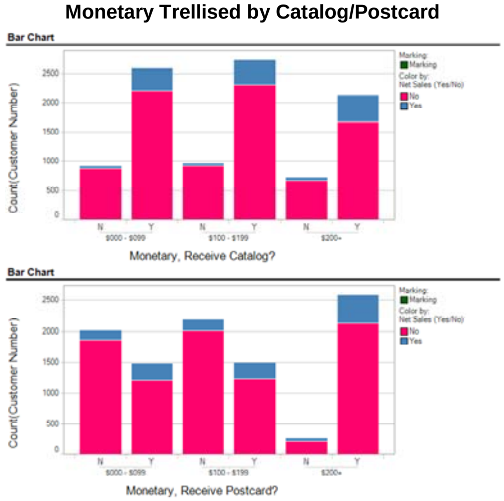
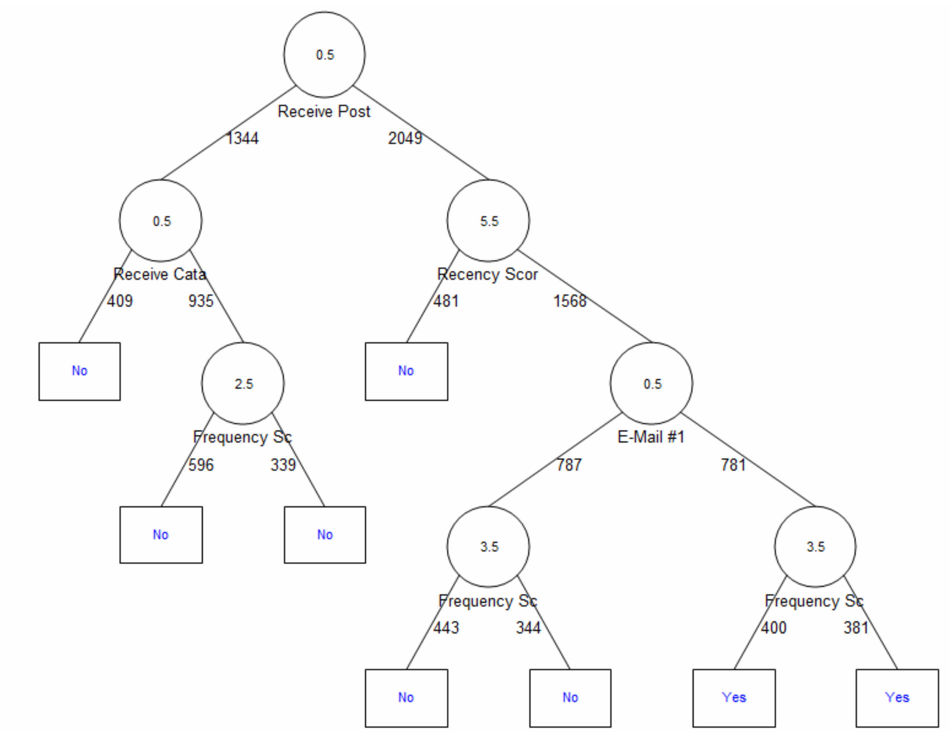


Exhibit 4 – Past Consumer Behavior Scoring

Recency	Recency Score	Frequency	Frequency Score	Monetary	Monetary Score
00-03 Mo.	7	8x+	4	$200+	3
04-06 Mo.	6	5-7x	3	$100 - $19	2
07-09 Mo.	5	2-4x	2	$000 - $09	1
10-12 Mo.	4	1x	1		
13-18 Mo.	3				
19-24 Mo.	2				
25-99 Mo.	1				

Exhibit 5 – Classification Tree Output (Best Model)



Classification Confusion Matrix		
	Predicted Class	
Actual Class	Yes	No
Yes	408	949
No	373	1663

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Yes	No
Yes	408	949
No	373	1663

Exhibit 6 – Logistic Regression Output (Best Model)

Input variables	Coefficient	Std. Error	p-value	Odds
Constant term	-1.9946413	0.13340604	0	*
Recency_07-09 Mo.	-0.3065798	0.10445041	0.00333364	0.73595977
Recency_10-12 Mo.	-0.51442212	0.12670441	0.00004907	0.59784597
Recency_13-18 Mo.	-0.53711236	0.18673055	0.00402235	0.58443344
Recency_19-24 Mo.	-0.46823183	0.23157243	0.04317975	0.62610835
Monetary_$200+	0.30286267	0.08044361	0.00016661	1.35372853
Receive Catalog? (Y/N)_Y	1.22650313	0.12660852	0	3.40928674
Receive Postcard? (Y/N)_Y	0.50547439	0.08411637	0	1.65777171

Residual df	4062
Residual Dev.	4808.669922
% Success in training data	33.34152334
# Iterations used	9
Multiple R-squared	0.07198765

Cut off Prob.Val. for Success (Updatable)	0.5

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Yes	No
Yes	280	1077
No	306	2407

Error Report			
Class	# Cases	# Errors	% Error
Yes	1357	1077	79.37
No	2713	306	11.28
Overall	4070	1383	33.98

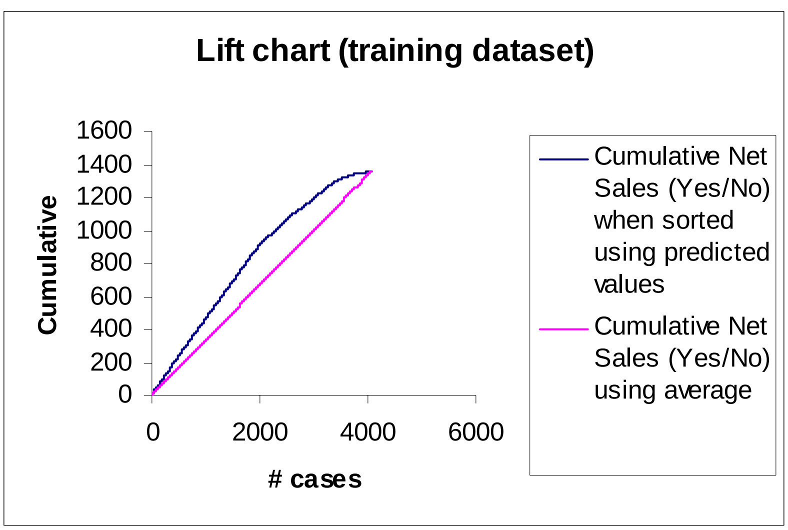


Exhibit 7 – Discriminant Analysis Output (Best Model)

Variables	Classification Function	
	1	0
Constant	-11.63763809	-10.43947792
Recency_04-06 Mo.	4.88352489	5.03142786
Recency_07-09 Mo.	5.72993994	6.11772013
Recency_10-12 Mo.	5.53590488	6.1128521
Recency_13-18 Mo.	13.45143223	13.891325
Recency_19-24 Mo.	15.7751503	16.13276291
Recency_25-99 Mo.	16.68216896	16.83932304
Frequency_2-4x	3.84875202	3.74452519
Frequency_5-7x	3.73601961	3.57422686
Frequency_8x+	2.57623243	2.33983088
Monetary_$100 - $199	3.0991044	3.02977943
Monetary_$200+	1.00279856	0.62376088
Receive Catalog?_1	10.65406513	9.65767765
Receive Postcard?_1	5.20571852	4.84912825
E-Mail #1 Score_1	-0.48178485	-0.585733
E-Mail #2 Score_1	2.92032552	2.88010263

Cut off Prob.Val. for Success (Updatable)	0.5

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	973	383
0	1231	1483

Error Report			
Class	# Cases	# Errors	% Error
1	1356	383	28.24
0	2714	1231	45.36
Overall	4070	1614	39.66

Exhibit 8 – Cost of Campaign – Return on Investment

	Unit Cost*	Total Cost
Catalog**	$0.78	$5,793.06
C. Postage	$1.51	$11,214.77
Postcard	$0.02	$88.69
P. Postage	$0.26	$1,441.18
Email 1	$0.01	$49.93
Email 2	$0.01	$39.81
Total Revenue	$ 131,589.93	$ 18,627.44
Advert/Sales	0.1416	

*We assume that the dataset we have is a snapshot of a larger campaign and are therefore basing costs on 100,000 units of each.

**Catalog cost based on 80# gloss, 4-color 40 page