Trojan Horse: Unexpected Style at Your Door

Trojan Horse is a (fictional) personalized fashion service for men. Similar to Trunk Club, BirchBox, and StitchFix, Trojan Horse provides its members with an assortment of high-end fashion accessories by mail that they can choose to purchase. To quote Brian Splay, the CEO of Trunk Club and Trojan Horse's main competitor, "Guys hate shopping but want to look good." Personalized fashion services take the hassle out of shopping.

Current Business Model

The business model is simple. Each month, Trojan Horse ships a box of men's fashion accessories to a portion of its members. Typically, they aim to ship to 50,000 members of their total 500,000 membership (10%). Members have 10 days to review the contents. If they like the accessories, they can purchase the box. If they do not, they can return the box at no cost—shipping and handling is free. Returned items are sold on Trojan Horse's website, or ultimately sold to third-party discount competitors at a deep discount.

An important feature of the business is that Trojan Horse does not manufacture any of its products; they purchase the products retail. Thus, if members wanted, they could purchase the same products themselves, without using Trojan Horse, slightly cheaper from the retailer directly. The reason members are willing to pay a premium is for the convenience of having accessories that are handpicked by Trojan Horse's stylists delivered directly to their door. Thus, it is critical for each promotion that Trojan Horse only ships to members that are likely to purchase the box.

Currently, Trojan Horse decides which members to choose for a particular box by identifying those members that have purchased similar items. Specifically, every box is classified by the stylists into one of 10 categories (See Exhibit 1) based on what kind of customers the stylists believe might purchase this type of box. Trojan Horse then ships to members who frequently purchase boxes from this category. For example, if this month's box is in the category "Hipster Chic," Trojan Horse scans its database for members that previously bought items from "Hipster Chic" and ships to them. The intuition is that customers each have their own individual style and are likely to purchase products that align with that style. The method is reasonably effective; on average 10% of members ultimately choose to buy the box.

A New Data-Driven Approach

You've been hired to explore the possibility of using more sophisticated business analytics techniques to choose which members to ship to for each promotion. The idea is that for each new box you first ship to a random set of 2,000 members. After two weeks, you will know which of those 2,000 customers chose to purchase the box. You can then use the Trojan Horse database to try and build a classification to model to predict which types of customers are likely to ultimately purchase the box (if shipped to them.) Using

this model, you can then target the most likely to purchase customer in the coming campaign.

Your job is to assess the feasibility and relative value of this new approach using the attached datasets. How much might the new approach increase profits? What other issues might arise from changing the shipment strategy? If you can suitably prove proof-of-concept with a small dataset, Trojan Horse's management is willing to invest in a full-scaled analytics team to do the implementation.

Sample Cost Data

For the case, you make the simplifying assumptions about the cost and revenue of a typical box:

Selling price per box	$60
Cost per box (retail)	$10
Salvage Value per box	$15
Cost of Mailing per box	$4.5
Cost of Returning per box	$4.5

Notice that the return-cost is only paid if the member decides not to buy the box.

Exhibit 1: The database includes the following information for each member:

Variable	Description
Gender	0 = Male 1 = Female
M	Monetary—total money spent at TrojanHorse in past 5 years
R	Recency—months since last purchase
F	Frequency—total number of purchases
FirstPurch	Months since first purchase
Surfer	Number of purchases from this category in past 5 years
BusinessExecutive	
Yuppie	
Hipster	
Artist	
ClassicGentleman	
Rugged	
Formal	
Casual	
Comic	

Based on the Training Data Set,

Level	Count	Prob
Non Buyer	888	0.88800
Buyer	112	0.11200
Total	1000	1.00000

The profit and cost per unit of product mailed is as follow.

Profit per Product after mailing cost	60-10-4.5= 45.5
Cost of Mailing the Product to not a buyer	-10-4.5-4.5+15 = -4

Suppose the company would like to select 50,000 potential buyers from its members for the next marketing campaign. The expected profit from these 50,000 potential buyers depends on the probability of purchase once they receive the products. The following calculations estimate the best and baseline profits.

(1)   **Maximum profit** (all customers will purchase)
The company can perfectly identify the buyers. In other words, all 50,000 customers who receive the product will purchase. The profit is 50,000 * 45.5 = $**2.275M**.

(2)   **Baseline profit** (no model, randomly select)
The **Baseline profit** if the Products are mailed randomly
 = 50,000 * 0.112 * 45.5 + 50,000 *0.888* (-4) = $**77,200**

The overall task of this case is to build a decision tree model that will help identify potential buyers and increase the expected profit of the campaign. Case questions and submission instructions are provided in the case question document.