

## Direct Mail Prospect Scoring Exercise

The goal of this assignment is to mimic the classic scoring activity that is regularly carried out for customer acquisition by firms engaging in one-to-one marketing. The firm in question is the Melrose Chocolate House. It has purchased two lists from the same list owner, one with 200 names and another with 300 names. When rented, the list is populated with 8 variables: (1) gender of the head of household, (2) frequency of visiting food related web sites, (3) frequency of visiting travel related web sites, (4) expenditure in the restaurant category, (5) expenditure in the travel category, (6) expenditure in the entertainment category, (7) income, (8), ethnic diversity of neighborhood.

All persons on both lists (500 in total) were contacted with a solicitation which contains (a) material introducing the Melrose Chocolate House, (b) a brief product listing, and (c) a coupon code for a deep introductory discount for a 1 pound box of Belgian-style pralines. We observe whether or not each of the 500 persons accepted this introductory offer and placed a purchase order: the value of  $y$  is 1 if the person placed an order and the value is 0 otherwise. The first list of 200 persons is used to estimate a scoring model. The second list of 300 is used to test the scoring model obtained from the first list, examine the distribution of prospects along various metrics and evaluate how successful the target selection was. The 200-person list will henceforth be referred to as the estimation-list, and the 300-person list will be referred to as the holdout list.

The list data are available from the class web page for this HW as an excel file; the file contains the data from *both* lists: the data from the estimation list first, followed by the data from the validation list.

This homework assignment entails following the steps below. Your written submission should consist *only* of those deliverables marked indicated by “Turn-in”.

1. Run a logistic regression on the estimation list predicting  $y$  as a function of gender and the seven hotline variables and the intercept (constant) term. Do not eliminate coefficients which appear statistically insignificant. Turn-in: The score equation for  $t$  from the logistic regression.

2. Evaluate the score function for each person in the holdout list. Based on this, compute the following for each person: (1) the predicted response probability, (2) the consequent lift, which is obtained by dividing the predicted response rate by the average response rate *in the estimation list*. No Turn-in.
3. Sort all the persons in the holdout list in decreasing order of lift as is done in slide 30. No Turn-in.
4. Plot the curve for marginal response rate vs number of solicitations made, like in the chart that appears on slide 31. Turn-in: the chart.
5. Suppose average LT Customer equity is \$30, and the solicitation cost is \$12. Use the marginal cost rule to calculate how many of the best prospects in the holdout list Melrose should send contact. Turn-in: your calculations and final answer.
6. Compute the cumulative sum (aka running sum) for the predicted response rates in decreasing order. Plot the curve for curve for number of positive responses vs number of solicitations made, like in the chart that appears on slide 33. You will notice that the curve does not rise as fast as one would expect from the usual 80/20 rule. Explain why the curve might be of this shape here. Turn-in: the chart and your explanation.
7. Assume that the firm has enough raw Belgian chocolate to make only 40 boxes of the pralines offered as part of the introductory offer. Use the limited supply rule to calculate how many of the best prospects in the holdout list Melrose should send solicitations to. Turn-in: your calculations and final answer.
8. To test how well the scoring model predicts actual performance, Melrose sent the solicitation offer to all 300 persons on the holdout list. The response is available as the  $y$  variable in the holdout list data part of the spreadsheet. Compute the cumulative sum (aka running sum) for this  $y$  variable in decreasing order of predicted response rate. Plot the curve for curve for number of *actual* positive responses vs number of solicitations made. Superimpose on this the curve obtained in step 6 above to get a chart akin to slide 44 . Based on this chart, comment on the discrepancies between the actual response rates and the predicted response rates for the prospects. Comment also on the consequent impact on the decision made in step 7 above. Turn-in: the chart and your comments.