|  |  |
| --- | --- |
| C:\Users\gps329\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\2DD80572.tmp | Marketing Analytics I |
| Fall 2021 |

**HOMEWORK EXERCISE 2**

**This exercise is to be submitted to Canvas by 2:00 PM on Thursday Oct 21, 2021. Please submit your answers in an Excel file with the following naming convention:**

**“last\_first\_hw2” (for example, sonnier\_garrett\_hw2)**

**Your submission file should be neatly formatted and contain five tabs, one for each question.**

**This is an individual homework assignment but you may confer with your classmates.**

**In addition to this document there is an Excel workbook (Customer Scoring Data 2021.xlsx) containing the data necessary to complete the exercise posted on Canvas.**

**I would recommend a concise explanation of how you calculated any numbers you are asked to report. This enables us to better give partial credit in the event of a mistake. Good luck.**

**Customer Scoring at Orange Apron**

Orange Apron is a subscription based meal delivery service that provides subscribers with three meal kits per week for 52 weeks of the year. Orange Apron is considering a customer acquisition campaign and plans to run a field experiment to determine appropriate target customers. To implement the campaign, Orange Apron has rented a list containing information on 500 households. The list contains some information about the households captured in four variables. The first variable is a binary indicator of whether children are present in the household (1=yes, 0=no). The remaining variables are three “hotline” buying indices. Similar to a credit rating, these indices are variables computed by the list owner and represent different index variables that generally indicate positive or negative purchase interest (for different product categories, some indices are positively correlated with purchase interest while other indices are negatively correlated with purchase interest). In consultation with the list owner, Orange Apron has selected three hotline indices, h1, h2 and h3. Orange Apron’s hypothesis is that h1 is positively correlated with interest in a meal delivery service while h2 and h3 are negatively correlated with interest in a meal delivery service. In addition, Orange Apron feels the presence of children in the household may increase interest in the service.

Orange Apron has sent an invitation to all 500 names on the list to join service. The invitation offer includes a deep discount on three weeks of service. We observe whether or not each of the 500 consumers accepted the invitation: the value of *y* is 1 if the person joined the service and the value is 0 otherwise. We use a random sample of 244 persons as the estimation sample (i.e., we estimate the scoring model on this data). The second list of 256 is used to test list scoring and evaluate how successful the target selection was. The 244-person list will henceforth be referred to as the estimation-list, and the 256-person list will be referred to as the holdout-list.

The list data are available on Canvas in the HW2 folder (Customer Scoring Data 2021.xlsx). *This exercise closely follows the class notes. You might want to review the class notes and replicate them with the Student Data used in class before attempting the exercise.*

1. Use a logistic regression model to predict *y* (i.e., the decision to join the club) as a function of the available scoring variables (children in HH and hotline variables). Include an intercept term in your model. You may run this model with XLSTAT rather than writing your own algorithm. For the homework, **keep all coefficients** (i.e., do not eliminate coefficients which seem to be statistically insignificant). *Report the parameter estimates of your resulting score function from the logistic regression as well as the p-values of the parameter estimates. What is your assessment of Orange Apron’s hypotheses? For each hotline variable, what effect does a 1-unit increase in the index have on the % increase in the odds of joining the service?*
2. Using your estimates compute the score for all individuals in the ***holdout*** data. Using the predicted score, compute (for each individual): (a) the predicted response rate and (b) the resulting lift (divide the predicted response rate by the actual average response rate *in the estimation-list****)*.** For each holdout individual, also compute the marginal effect of a 1 unit change in each of the hotline variables on the response probabilities.

*Report the estimated score, response rate, lift, and marginal effects for the holdout id’s and report the* ***average*** *response rate, lift and marginal effects computed over all 256 holdout individuals.*

1. Sort the holdout-list in decreasing order of response probability (or equivalently, lift). Plot the expected and actual sales from sending N solicitations to the N best customers for N=1 to 256. *Report the resulting graph. What is your assessment of the model’s predictive performance?*
2. The grocery and meal delivery business is notorious for low margins and high customer churn. Orange Apron estimates the average customer lifetime value to be $13.50. Assume a solicitation cost of $3. Based on the marginal cost rule (see class notes) determine the cut-off probability for making offers. Assume the list owner charges a rental cost of $1 per household. What is the improvement in profits from targeting? *Report your calculations for determining the cut-off and the % of the holdout list customers that would be targeted. Report the profits from your targeting strategy. How does this compare to soliciting all 256 households?*
3. What is your assessment of the monetary value of the demographic information versus the three hotline variables (consider other models with all the hotline variables together as a set of information to be compared with the demographic information)?