# Customer Acquisition Scoring Model

# Usefulness

- The customer acquisition process and data can be a useful model to determine a customer's value and potential to buy before spending money on marketing efforts.
- This assignment highlights my knowledge in building a logistic regression model and reporting an estimated score, response rate, and lift rate.
- It also highlights the importance of using targeted marketing efforts to maximize profits. Through this analysis, I will prove that targeting is an essential marketing strategy in customer acquisition.

# Data Set

## Estimation Data

| id | children in HH | hl1 | hl2 | hl3 | y |
|---|---|---|---|---|---|
| 1 | 1 | 30 | 0 | 0 | 0 |
| 2 | 0 | 22 | 10 | 26 | 0 |
| 3 | 0 | 20 | 45 | 13 | 0 |
| 4 | 1 | 15 | 15 | 0 | 0 |
| 5 | 0 | 4 | 15 | 0 | 0 |
| 6 | 1 | 18 | 0 | 0 | 1 |
| 7 | 1 | 16 | 0 | 0 | 0 |
| 15 | 0 | 5 | 0 | 0 | 1 |
| 16 | 1 | 16 | 0 | 13 | 1 |
| 17 | 0 | 21 | 20 | 13 | 0 |
| 18 | 1 | 15 | 20 | 0 | 0 |
| 19 | 0 | 6 | 0 | 0 | 0 |
| 22 | 1 | 38 | 20 | 0 | 1 |
| 24 | 0 | 31 | 15 | 0 | 1 |
| 25 | 1 | 26 | 0 | 0 | 0 |
| 26 | 0 | 5 | 10 | 0 | 0 |
| 29 | 1 | 4 | 0 | 0 | 0 |
| 30 | 0 | 46 | 10 | 13 | 1 |
| 31 | 1 | 10 | 0 | 13 | 0 |

## Holdout Data

| id | children in HH | hl1 | hl2 | hl3 | y |
|---|---|---|---|---|---|
| 8 | 1 | 47 | 40 | 0 | 1 |
| 9 | 0 | 45 | 20 | 26 | 0 |
| 10 | 0 | 11 | 0 | 15 | 0 |
| 11 | 0 | 17 | 10 | 0 | 0 |
| 12 | 0 | 9 | 0 | 0 | 0 |
| 13 | 0 | 22 | 10 | 13 | 1 |
| 14 | 0 | 35 | 35 | 13 | 0 |
| 20 | 0 | 14 | 0 | 0 | 0 |
| 21 | 1 | 17 | 0 | 13 | 1 |
| 23 | 0 | 12 | 0 | 0 | 0 |
| 27 | 1 | 6 | 10 | 0 | 0 |
| 28 | 1 | 26 | 0 | 0 | 1 |
| 32 | 0 | 30 | 15 | 0 | 1 |
| 33 | 1 | 20 | 10 | 0 | 0 |
| 35 | 0 | 24 | 15 | 0 | 0 |
| 36 | 0 | 15 | 15 | 0 | 0 |
| 38 | 1 | 31 | 15 | 0 | 1 |
| 39 | 0 | 30 | 35 | 13 | 0 |
| 41 | 0 | 8 | 0 | 15 | 1 |

- The first variable is a binary indicator of whether children are present in the household (1=yes, 0=no).
- Orange Apron has selected three hotline indices, h1, h2 and h3.

"hotline" buying indices: Like 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).

# Hypothesis

- 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

# Analysis

First, I ran a logistic regression to predict the decision to buy the subscription service as a function of the available scoring variables.

| Model parameters (Variable y): | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Source | Value | Standard error | Wald Chi-Square | Pr > Chi² | Wald Lower bound (95%) | Wald Upper bound (95%) | Odds ratio | Odds ratio Lower bound (95%) | Odds ratio Upper bound (95%) |
| Intercept | -1.512 | 0.355 | 18.156 | <0.0001 | -2.208 | -0.817 | | | |
| children in HH | 0.904 | 0.301 | 9.037 | 0.003 | 0.315 | 1.493 | 2.469 | 1.370 | 4.450 |
| hl1 | 0.033 | 0.016 | 4.452 | 0.035 | 0.002 | 0.064 | 1.034 | 1.002 | 1.067 |
| hl2 | -0.027 | 0.012 | 4.915 | 0.027 | -0.051 | -0.003 | 0.973 | 0.950 | 0.997 |
| hl3 | -0.004 | 0.016 | 0.049 | 0.825 | -0.035 | 0.028 | 0.996 | 0.966 | 1.028 |

From this, I concluded that Orange Apron's hypothesis was mostly correct. Through the values calculated using the Logit Linear Regression, we can see that there is a positive correlation with hl1 (0.033) and negative correlations with hl2 and hl3 (-0.027, -0.004). However, looking at the p-values, <u>the hl3 value is not significant</u>. In addition, there is a strong positive correlation with children in HH (0.904), meaning that the presence of children does increase interest in the service.

The odds ratios help us understand what effect a 1-unit increase in the index has on the % increase in the odds of joining the service.

- hl1: With a 1-unit increase, there is a 3.4% increase in the odds of joining the service.
- hl2: With a 1-unit increase, there is a 2.7% decrease in the odds of joining the service.
- hl3: With a 1-unit increase, there is no effect as the coefficient is not statistically significantly different from zero

# Analysis

Next, I used my regression estimates to compute a score for each individual in the dataset and the average response rate, average lift rate, and marginal effect for each variable alone across the 256 individuals in the holdout data.

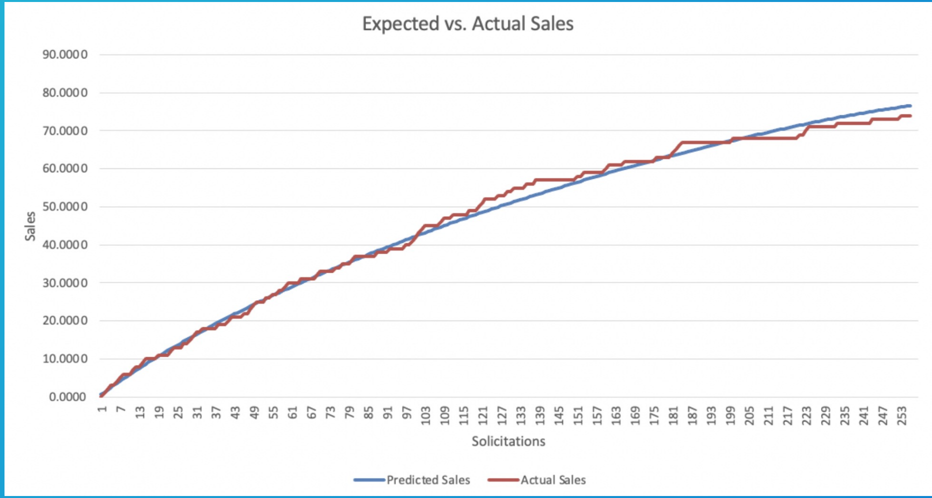
| Score | Prob | Lift | Marg Effect of Change in hl1 | Marg Effect of Change in hl2 | Marg Effect of Change in hl3 |
|---|---|---|---|---|---|
| -0.12461 | 0.468889 | 1.505379 | 0.008321 | -0.00676 | -0.00088 |
| -0.64436 | 0.344262 | 1.105261 | 0.007543 | -0.00613 | -0.0008 |
| -1.19814 | 0.231807 | 0.744221 | 0.00595 | -0.00484 | -0.00063 |
| -1.21595 | 0.22865 | 0.734088 | 0.005893 | -0.00479 | -0.00063 |
| -1.21166 | 0.229407 | 0.736519 | 0.005907 | -0.0048 | -0.00063 |
| -1.09508 | 0.250663 | 0.804761 | 0.006276 | -0.0051 | -0.00067 |
| -1.33969 | 0.207561 | 0.666379 | 0.005496 | -0.00447 | -0.00058 |
| -1.04459 | 0.260265 | 0.835587 | 0.006433 | -0.00523 | -0.00068 |
| -0.08683 | 0.478307 | 1.535617 | 0.008338 | -0.00678 | -0.00089 |
| -1.11142 | 0.247606 | 0.794947 | 0.006225 | -0.00506 | -0.00066 |
| -0.67977 | 0.336312 | 1.079738 | 0.007458 | -0.00606 | -0.00079 |
| 0.260094 | 0.564659 | 1.812854 | 0.008214 | -0.00668 | -0.00087 |
| -0.91737 | 0.285494 | 0.916587 | 0.006816 | -0.00554 | -0.00072 |
| -0.21198 | 0.447202 | 1.435752 | 0.00826 | -0.00671 | -0.00088 |

→ →

| Avg Resp Rate | Avg Lift | Avg. Effect hl1 | Avg. Effect hl2 | Avg. Effect hl3 |
|---|---|---|---|---|
| 29.903% | 96.004% | 0.652% | -0.530% | -0.069% |

# Analysis

My next step was to sort the holdout data list in decreasing order of response probability. I then plotted the expected and actual sales from sending N solicitations to the N best customers for N=1 to 256.




The model seems to predict actual sales reasonably well. Once we are through the ~100 best customers the model underpredicts actual sales then tends to over predict.

# Analysis

Knowing that the grocery and meal delivery business is notorious for low margins and high customer churn, I was curious to see what the improvements in profits would be from targeting only the customers that were above the cut-off response rate of 22%. I chose 22% because that was the breakeven rate.

| id | y | Prob | Target | | | | | Profit |
|---|---|---|---|---|---|---|---|---|
| 8 | 1 | 0.468889 | 1 | | ECLV | 13.50 | | 9.50 |
| 9 | 0 | 0.344262 | 1 | | Solicitation | 3 | | -4.00 |
| 10 | 0 | 0.231807 | 1 | | Cutoff | 0.222222 | | -4.00 |
| 11 | 0 | 0.22865 | 1 | | Rental | 1 | | -4.00 |
| 12 | 0 | 0.229407 | 1 | | | | | -4.00 |
| 13 | 1 | 0.250663 | 1 | | % Targeted | 0.691406 | | 9.50 |
| 14 | 0 | 0.207561 | 0 | | | | | -1.00 |

| Profits with Targeting | Profits Without |
|---|---|
| 63.50 | -25.00 |

# Results

- Through this analysis, I proved that targeting was an effective strategy when considering marketing efforts for customer acquisition as the profit was higher.
- I also showed how different variables that play into a customer's potential to join the business can affect the targeting decisions.