**Organic Product Case Study**

***Business Scenario***: A supermarket is offering a new line of organic products. The supermarket’s management wants to determine which customers are likely to purchase these products. They are also interested in understanding the profitability of their customers as well. If they notice that customers who purchase organic products are also highly profitable customers, that makes the additional cost of stocking organic products more palatable for the management.

The supermarket has a customer loyalty program. As an initial buyer incentive plan, the supermarket provided coupons for the organic products to all of the loyalty program participants and collected data that includes whether these customers purchased any of the organic products.

The ORGANICS data set contains 13 variables and over 22,000 observations. The variables in the data set are shown below with the appropriate roles and levels:

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Model Role** | **Measurement Level** | **Description** |
| ID | ID | Nominal | Customer loyalty identification number |
| DemAffl | Input | Interval | Affluence grade on a scale from 1 to 30 |
| DemAge | Input | Interval | Age, in years |
| DemCluster | Rejected | Nominal | Type of residential neighborhood |
| DemClusterGroup | Input | Nominal | Neighborhood group |
| DemGender | Input | Nominal | M = male, F = female, U = unknown |
| DemRegion | Input | Nominal | Geographic region |
| DemTVReg | Input | Nominal | Television region |
| PromClass | Input | Nominal | Loyalty status: tin, silver, gold, or platinum |
| PromSpend | Input | Interval | Total amount spent in the store this year |
| PromTime | Input | Interval | Time as loyalty card member |
| TargetBuy | Target | Binary | Organics purchased? 1 = Yes, 0 = No |
| TargetAmt | Rejected\*\* | Interval\*\* | Number of organic products purchased\*\*  Note that you could use this as an input for exploratory analysis – You just don’t need to consider it as a Target variable for this exercise (which is why it was rejected in the original analysis of these data) |

**Tasks and Questions**:

You have two major analytical tasks – (1) Explore the data in order to better understand the profitability of customers who purchased organic products vs those who didn’t and (2) build a predictive model for classifying customers according to their likelihood to purchase organic products. You can use logistic regression or decision trees or you can try both methods to see which performs the best. You can use JMP or Enterprise Miner or Excel or any combination of the 3 products to help you explore the data, build the models, and answer the questions. I have provided the data as both a SAS data set and an Excel file. If you use JMP for your analysis, it can read both of those file formats directly.

Note that there are some missing values in this data set, so be sure to have a strategy for dealing with those.

You will need to submit the answers to the following questions (some of which will require you to use the output from your predictive model). Note that you can should use PromSpend as a proxy for “profitability” in the questions below:

1. This first two questions are purely exploratory analysis questions. How would you characterize the “profitability” of the customers who purchased organic products vs those who didn’t purchase organic products? Do they spend similar amounts, or does there appear to be a significant difference? Do customers who purchase organic products spend more at your store in general than customers how don’t purchase organic products (or vice versa)?

Distribution TargetBuy

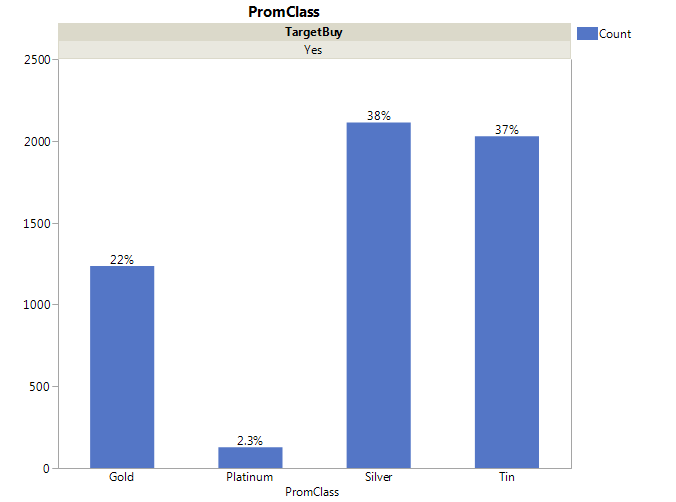
No > Yes

Distribution -> PromSpend, Y by TargetBuy

Conclude: they spend similar amounts (The No group is a little bit higher)

Question to ask: what does the last question mean (spend more time or more $?) ? What “characterize the profitability” means?

1. Continuing along a similar path, are there any noticeable differences in the percentage of customers who purchase organic products across the different loyalty status groups (for example, is the percentage of platinum customers who purchase organic products higher than the percentage of tin customers who purchase organic products)? What about the profitability of the customers in the different loyalty groups?



Graph 2a: Percentage of customers who purchase organic products across different loyalty status group

Question to ask: what the last question means?

1. What factors seem to have the most impact on a customer’s likelihood to purchase organic products? (include any relevant statistical output to support your answer) Based on your model, how would you describe the “typical” organic products customer?

Age, gender, neighborhood group, geographic group, TV group, Loyalty Time

1. As a result of this modeling exercise, what strategy would you suggest to the supermarket if they are interested in trying to grow their organic product line while continuing to retain profitable customers and grow their bottom line? Your recommendation about how you would do with shopping. Promotions: coupon campaigns, loyalty rewards points and stuff.
2. Five new customers signed up for the loyalty program this month. Three of them have recently relocated to the area and have been a part of the loyalty program for the supermarket chain in another state. Their information has been saved in the Excel file “SuperCustomers.xlsx.” For each of these customers, determine whether they are a good candidate for a promotion related to organic products (e.g., are they likely to purchase organic products or not).

Take these customers drop into the decision trees

Are they likely to buy the Organics or not

Promspend grouped by Target amount

Promspend (Y), promclass (bottom), Target (top)