Boats (B):

A Segmentation Case

This case was written by Joerg Niessing, Associate Partner and Head of Analytics at Prophet, and Theodoros Evgeniou, Associate Professor of Decision Sciences and Technology Management at INSEAD. It is intended to be used as a basis for class discussion rather than to illustrate either effective or ineffective handling of an administrative situation. All numbers in this case are not real.

A few weeks later Mary and her team had a successful strategy meeting with the CMO and the CEO where Mary presented the different segments that she and her team identified. Everybody got a good understanding about the different segments but for one important question Mary had no good answer so far.

The Business Problem

Mark Sundowner, the CMO of CreeqBoat, asked Mary if it is possible to get more insights around the key purchase drivers for each segment. The profiling of the segments she presented was really crisp and the board had a good understanding about the typical person that describes each segment in terms of their attitudes and their demographics. But Mark wanted to get a better understanding on the "WHY" – why do people in these segments purchase a boat? What are the key elements that are driving a future boat purchase? Mark wanted to sharpen the positioning of CreeqBoat and make it more relevant for their target segments. It would also help him to come up with more tailored and targeted marketing strategies for the key segments. He was hoping that this purchase driver analysis would generate the insights needed.

The Data

Mary recently was visiting a conference about "Customer insights and Marketing analytics" where she heard about the power of a data-derived analysis using logistic regression models. The data-derived analysis was known to be much better than a simple stated importance question about potential purchase drivers one could directly ask customers.

Mary and her team took another detailed look into the data they gathered about the boating market in the US. Luckily she had included several more questions that allowed her and her insights team to perform a data-derived purchase driver analysis:

1. In the research she included a question where she asked each respondent for *one* brand whether they would consider purchasing that brand in the future (Q. 18 in the Appendix). The brand was selected so that the respondents were familiar with the brand shown in that question.

2. In addition she asked each respondent almost 27 brand equity ratings for the *same* brand (Q.16, 1-27). The respondents needed to rate the brand on various statements that describe product, image or reputation of that brand (Q.16 in the Appendix)

The Real Work

Using a logistic regression with the 'future purchase' (Q. 18 in the Appendix) variable as the dependent variable she could understand in a derived way 'WHY' people would purchase that brand. Mary was so excited that she started with the analysis right away. She wanted to find answers for the following questions:

1. What are they key purchase drivers in the boating industries overall across all segments?

2. What are they key purchase drivers for each individual segment? Are the drivers that different that every segment needs to be targeted individually? What criteria would you consider in deciding which segment to focus on, and why?

3. What would be the recommended strategy without segmentation? How about the recommended strategy with taking into account segmentation? Are the strategies different? In what way?

Always being sceptical and careful, Mary was also worried about the reliability of her results and wanted to test whether a different approach would lead to similar decisions. She decided to examine the key purchase drivers for the largest segment using another data analytics approach. In the past her team had used *Classification Trees* to understand what differentiates people interested in their products from the rest. Remembering the intuitive feel of the results of such trees, she decided to do the drivers analysis also using this method.

1. Use Classification Trees to discriminate future purchase intenders and find what are the main drivers that separate them from the non-intenders for the largest segment. What are the key purchase drivers found with this approach? How do they compare intuitively with the drivers Mary found using logistic regression?

2. For the largest segment, statistically compare the two approaches as follows. First split that segment into a training and a validation set (e.g. 75-25% split). Then develop both a logistic regression and a Classification Tree model using the training set only. Finally, for both the estimated models, generate a misclassification table, a lift curve, and an ROC curve using both the training and the validation sets. How do the two approaches compare? What if you consider different Classification Tree models?

(Optional) While doing the analysis Mary also realized that she not only added a question around 'future purchase', but also she asked all respondents if they would recommend the brand they rated to a friend (Q. 17 in the Appendix). She wanted to impress Mark by understanding the 'drivers of recommendation' as well by using that variable as a dependent variable.

3. What are they key 'drivers of recommendation' in the boating industries for the largest segment Mary found? How does it differ from the 'purchase drivers' for that segment? For simplicity Mary decided to use again logistic regression for this analysis.