Predict 450 Marketing Analytics Section 55

Solo 3: XYZ Company Database Campaign Response Modeling

Your Mission

Use XYZ's data to develop a model-based method for targeting customers that XYZ can use to make its upcoming mail campaign profitable. This campaign will consist of one mailing to XYZ customers who are selected based on your analytic results and your recommendations.

1. Develop the best predictive model you can using initial responses (i.e., binary responses to first mailing) to XYZ's most recent mail campaign (the 16th one) as your dependent variable. You may consider and apply either Frequentist or Bayesian predictive modeling methods, or both. Assess different models, and compare them using statistical methods to select your best one.

   A "good" model will have predictive accuracy, and the more of it, the better. It will help inform XYZ about which of its customers to mail to. It might even suggest things that XYZ can to increase the likelihood of sales by changing its marketing mix.

   Assume that it costs XYZ $3.00 to produce and mail a catalogue to a customer, and that XYZ has recently earned a 10% profit on items it sells through its catalogues, excluding the costs of the catalogue and mailing.

   Using your model results and the other information available, identify which of XYZ's customers should be sent a catalogue, and what net revenue (net profit from sales less mailing costs) you expect will result from mailing to them. Include in your report how you did your calculations.

2. "Score" (see below) customers that XYZ didn't target in its most recent campaign so that XYZ can mail to those you predict will produce net revenue as a result. Describe your method and include your estimate of expected net revenue from mailing to these customers. Be sure to mention any key assumptions or limitations.

3. Compare the net revenue you expect based on your modeling to the revenue that actually resulted from XYZ's 16th campaign, and explain any difference you observe.

   a. Interpret the results of your most preferred model from a marketing perspective so that XYZ can learn about its customers.

   b. Address how XYZ can apply your final model to new customers. XYZ would like to be able to predict responses to their next campaign by customers who are yet to be acquired by XYZ.

   c. Suppose that it would be possible to assess your model's ability to predict responses to XYZ's next (The 17th) campaign by targeting selected customers. How would you design this "test?" What comparisons would you make?

XYZ is counting on you to help make their next campaign as lucrative as possible.

Be sure to succinctly summarize your methodology, and also the basis(es) for the various decisions you made, and the conclusions you drew. Be sure to address all objectives and issues, and to qualify your results by describing any limitations of your methodology.

Be sure to see **Guidance**, below.

Materials

This assignment involves working with XYZ's database, which is more complicated than the data sets you used previously in this course. On Canvas you'll find some R data files, some documents describing the data, and some supporting materials. There are also videos about the creating the analytic data set and about evaluating targeting models in the Connect Recordings.

Working with the Data

Although there is documentation on XYZ's data, it is neither complete nor completely correct in all cases. You'll need to make the best use of the available information about the data that you can, as you would if you were doing a "real" project.

Prof. Miller has done much of the slog work in putting an analytic data set together for you. To fully understand the variables he has created for your use, you should review his R code, which although lengthy is pretty well documented. You'll find it in the file miller_xyz_data _integration.r. Open this file in an editor and review his code and comments. You'll also want to be sure to view the Connect recording he made describing how he constructed this combined data set. The recording is called "Special Solitary Sync Session to Review XYZ data work."

If you happen to take PRED 420, you'll run into this XYZ data, or at least versions of it, again.

Guidance

This is a target marketing response modeling exercise. Your task is to develop a model that predicts responses to the last (the 16th) mail campaign using the data you have available, and then to use it to tell XYZ which customers it should send a catalogue to in its next (17th) mail campaign so that the net returns (total returns less mailing costs) realized will be as large as possible.

You can use any of a number of different kinds of classifier or binary limited dependent variable regression models to do this kind of response modeling. For summarizing models and comparing them, you'll want to consider using ROC curves and lift charts, so see R packages like gains, ROCR, AUC, Epi, and BCA. (A question you might want to consider is, how can you tell whether the AUCs of two models differ reliably, i.e. "significantly?" How about whether an AUC is reliably greater than 0.5?)

Many would start out on on a task like this by fitting a binary logistic or probit regression model. Either may suffice for this assignment if its performance is adequate. But you could (also) use other methods as well, like cluster-wise regression or CART, an ensemble technique like Random Forest, or a margin classifier like SVM. You can use Frequentist methods or Bayesian methods. Or both. It's your choice. Your task is to produce the most useful model you can, one that will help XYZ get the

greatest returns from its next campaign.

About the Data:

There are three R binary data files available:

XYZ_customer_data.RData
XYZ_item_data.RData
XYZ_mail_data.RData

A simple R program for reading in the data is provided in the R source code file:

XYZ_read_data.r

And, as you know, and as your luck would have it, an analytic data file has been created for your use:

XYZ_complete_customer_data_frame.RData

The code used to create this file is in the following file. *Be sure to review it and the video describing its construction so that you understand the variables in it:*

miller_xyz_data_integration.r

Data dictionaries are provided in an Excel file (XYZ_Data Dictionaries.xls) and also an Adobe Acrobat pdf file (XYZ_Data Dictionaries.pdf). Here's what should be a useful tip: when modeling responses, make sure you are using data from customers who were actually targeted by a campaign. Here's another: don't fall into the trap of predicting customer responses using a variable that's another way of measuring the same response.

Format

Last but not least, be sure to follow the formatting and submission instructions provided on Canvas for this assignment. Post your questions to the Solo 3 Huddle. Now it's up to you to see that XYZ Goes Forth and Prospers.