Wine Retailer Case

You are working for a wine retailer as a data analyst. You recently joined the team that is running email marketing experiments to evaluate offers prior to sending the offers to a broader set of customers. Your understanding is that these tests are well-structured experiments. You heard about a previous experiment involving two different emails and a control, and even saw some related code.

In this experiment they were examining the impact of an email that was intended to drive purchases. They want you to evaluate whether the promotion is effective and who to target with the email campaign. Although there is no direct cost to sending the emails, they recognize that fatigue among their email list results in dropped emails or lower responsiveness. Although this idea is an ongoing project to develop individual-level estimates of likely fatigue, currently the analytics team estimates the average cost of sending an email of approximately 10 cents. Also, while your margins vary on products the team has been using a simple 30% margin rule to apply to all products. The management team is quite sophisticated with statistics and familiar with this setting.

Your task is to analyze the data and provide

a) An evaluation of the average causal effect to be inserted in the main presentation.
b) Slice and dice analysis to illustrate the potential for targeting on responses for this email campaign. You should identify a good example of how slicing and dicing the data can lead to significant and economically meaningful differences in causal effects.
c) "Individual-level" conditional causal effect estimates to be used along with the margin and cost information to score each customer. Use the best approach to scoring available. Make predictions for the training sample and save these to a file. This file should not be included in your submission, but relevant summary information should be included in the presentation such as the percentage of training data that is targeted and summaries of the baseline variables for the targeted and non-targeted groups.

You should submit two files:

1) Powerpoint presentation with up to 5 slides beyond the title slide.
    a. Executive summary/highlights slide
    b. Methodology slide that indicates the key method for forming the score and targeting rules.
    c. Detail slides (2-3) providing details of the average causal effect, the slice and dice illustrations, and summary information about the targeted customer set
2) .Rmd file that contains all code to create the presentation, generate scores and targeting decisions. This file should contain a statement that writes a .csv file containing customer identifiers, score, and targeting indicator. The approach used here should be easily extended to predict on new data. The code should illustrate this with a commented out line(s) that assume a data.frame with "newdata" is available and makes the score and targeting decisions.

Evaluation

Frame (2 pts):
- Is the presentation complete and executes the requirements as specified?
- Are the best methods selected to perform the analysis?

Analyze (6 pts):
- Are estimates accurate?
- Is uncertainty properly represented?
- Is the slicing and dicing illustration effective?
- Is code for targeting and scoring correct?
- Is code for predicting on newdata correct and clearly represented?

Communicate (2 pts):
- Does the .Rmd file follow the required assignment format?
- Does the presentation tell the story of the data?
- Are insights/recommendations/analysis easy to locate (i.e., on each slide)?
- Are the tables/figures easily understood without explanation?
- Are the sources and inputs to figures clear and correct?
- Is the language and level of detail appropriate for a managerial audience?