**Week 7: Case study Assignment**

**21.7 Direct-Mail Fundraising**

**EVERY PART IN THE ASSIGNMENT MUST BE DONE IN R MARKDOWN AND *KNIT the RMD FILE TO PDF***

***Keep Author Name As Wenkkatessh***

***In the last You need to send RMD file and PDF.***

(TEXTBOOK page: 493) I Will also send the text book with this word file

Step 1: Exploratory Data Analysis (EDA) and Data Visualization. This step includes visualization and preliminary analysis of the  
dataset in the case. Your submitted report must include descriptive statistics, visualization, and commentary describing the key variables in the case along with the code.

Step 2: Methodology: Step 2 involves a description  
of the methods used for the case

Step 3: Complete Case: The last step of the  
assignment need to answer Assigned  
case questions. (See Page 4 of this word document for questions)

**Format of the Assignment must be**

*As a style reference, consider*[*this Kaggle.com competition*](https://nam10.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.kaggle.com%2Fcode%2Ferikbruin%2Fhouse-prices-lasso-xgboost-and-a-detailed-eda&data=05%7C01%7Cwenkkatessh.peraom%40scranton.edu%7Cf3c65cd78cb6439d9afe08da5308f02b%7Ca8edc49a41f14c699768a7f6d7c3b8c3%7C0%7C0%7C637913595097749467%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=VDVXolv1vOHrly%2FxKq4sr3kCEljwTDM1C6hlerm646c%3D&reserved=0)*;  it is organized,  informative, and to the point*.

**Description about Case Study Assignment**

***21.7 Direct-Mail Fundraising***

Fundraising.csv and FutureFundraising.csv are the datasets used for this case study. Background

Note: Be sure to read the information about oversampling and adjustment in Chapter 5 before starting to work on this case.

A national veterans’ organization wishes to develop a predictive model to improve the cost-effectiveness of their direct marketing campaign. The organization, with its in-house database of over 13 million donors, is one of the largest direct-mail fundraisers in the United States. According to their recent mailing records, the overall response rate is 5.1%. Out of those who responded (donated), the average donation is $13.00. Each mailing, which includes a gift of personalized address labels and assortments of cards and envelopes, costs $0.68 to produce and send. Using these facts, we take a sample of this dataset to develop a classifi- cation model that can effectively capture donors so that the expected net profit is

maximized. Weighted sampling is used, under-representing the non-responders so that the sample has equal numbers of donors and non-donors.

***Data***

The file Fundraising.csv contains 3120 records with 50% donors (TAR- GET\_B = 1) and 50% non-donors (TARGET\_B = 0). The amount of donation (TARGET\_D) is also included but is not used in this case. The descriptions for the 22 variables (including two target variables) are listed in Table 21.9.

Table

Description automatically generated

ASSIGNMENT QUESTIONS FOR STEP -3 is below

Text, letter

Description automatically generated

Note:

* Steps 2 and 3 of the assignment, the student must independently learn and apply to the problem in the case one data analytic method that is not covered in class.
* The case in the textbook asks you to fit several predictive classification models of your choice (see question 1 of the case).
* Let’s agree to fit three different models: **logistic regression, classification trees,**and**neural networks**. The logistic regression method is the subject of Chapter 10, the classification tree method is the subject of Chapter 9, and the neural networks method is the subject of Chapter 11 of the course textbook (Schmueli et al.)