Assignment 3

Use the Dataset 2 and keep the data of `app_key`=80862143 (Hero) for this assignment.

Q1. Provide the summary of the data. You can submit the page 18 of Lecture Note 5 after filling in all the blanked items of the page.

Q2. Generate a new variable, `is_dolphin`, where 1 if total expenditure>= 100,000 and 0 otherwise. Generate usage variables on page 19 of Lecture Note 5. Also, use your imagination to develop other variables which could be correlated to be high spenders.

Q3. Develop the best classification model which correctly identifies the dolphin users. Employ tree models and logit models. Choose the model and explain why.

Note: Your score can be affected by the level of sensitivity and specificity of your model.

Guideline for Assignment 3:

Submit your answer and R-codes used for the analysis to the course Blackboard. Please include your student number and name in the header of the document. Your document should be in the same format as the final report (See the syllabus) and should not exceed three A4 pages. Append R-codes used for the analysis. There is no page limit for the appendix. The due is Apr 15, 23:59PM.