Assignment for

Creative Gaming: Uplift Modeling

You decide to try uplift modeling to pick consumers for Zalon. You plan to use the experimental control group ("**Group 1**"), i.e., the 30,000 customers who did <u>not</u> receive an ad), as well as "**Group 2**", i.e., the randomly chosen experimental group of 30,000 customers who <u>did</u> receive the ad.

Part 1: Uplift Modeling Using Machine Learning (Random Forests)

1. Prepare your data:

   Hint: Please visualize what you are doing by looking at the new data frames you create in each step.

   a) Create "Group 2" by sampling from cg_ad_treatment. You can use this syntax:

   ```
   cg_ad_random <- cg_ad_treatment[sample_random_30000,]
   ```

   b) Create a stacked dataset for the uplift analysis by combining cg_organic_control (Group 1) and cg_ad_random (Group 2). You can use this syntax:

   ```
   expdata_stacked <-
      rbind(cg_organic_control %>% mutate(ad = 0),
            cg_ad_random %>% mutate(ad = 1))
   ```

   c) Split the stacked dataset into a training and test dataset. You can use this syntax:

   ```
   set.seed(1234)
   split.index <- stratified(expdata_stacked, c("ad",
                  "converted"), 0.7, bothSets=TRUE)
   expdata_stacked.train <- split.index[[1]]
   expdata_stacked.test <- split.index[[2]]
   ```

Professor Florian Zettelmeyer prepared this case to provide material for class discussion rather than to illustrate either effective or ineffective handling of a business situation. Events and people in this case are fictionalized.

2. Train an uplift model using random forests. Add the predicted scores for the treatment and control models to `expdata_stacked.test` and calculate the uplift score.

    Hint: Please see the handout "Random Forests in R.pdf" for how to run a random forest.

3. Calculate the Uplift (%) and Incremental Uplift (%) for the <u>uplift</u> model (use 20 instead of the standard 10 groups) and plot performance metrics. Interpret the plots.

4. Using the `incremental_Y1` column from the performance metric table created by `QiniTable()`, calculate the incremental profit you expect to make if you targeted the best 30,000 consumers of 120,000 using the <u>uplift</u> model.

    Hint: For every n-tile, the `incremental_Y1` tells you how many incremental purchases were made when consumers up to that n-tile were targeted. To extrapolate correctly to picking the best 30,000 from 120,000, notice that there are a total of 9,000 consumers who got the ad in the test sample `expdata_stacked.test`.

5. Calculate the Uplift (%) and Incremental Uplift (%) you would get if you used a <u>propensity</u> model (use 20 instead of the standard 10 groups). Compare the Uplift (%) performance metric between the uplift and propensity models. Interpret the difference.

    Hint: To compare the performance of the uplift and propensity models, use the functions `QiniCurve()` and `QiniBarPlot()`

6. Using the `incremental_Y1` column from the performance metric table created by `QiniTable()` for the <u>propensity</u> model, calculate the incremental profit you expect to make if you targeted the best 30,000 consumers of 120,000 using the <u>propensity</u> model. How much more money do you expect to make from using an <u>uplift</u> instead of a <u>propensity</u> model?

Part 2: Targeting the optimal percent of customers

So far we have always targeted a 25% of model-selected customers (by picking the best 30,000 out of 120,000 customers). We now want to evaluate whether we should target fewer or more than 25% of customers.

1. What formula would you use to select which consumers to target using a <u>propensity</u> model where your goal is to maximize <u>profits</u>? What percentage of customers in the ad treatment group of `expdata_stacked.test` would you target using the <u>propensity</u> model?

2. What formula would you use to select which consumers to target using an <u>uplift</u> model where your goal is to maximize <u>incremental profits</u>. What percentage of

customers in the ad treatment group of `expdata_stacked.test` would you target using the <u>uplift</u> model?

3. Rounding the targeting percentage numbers you calculated in 1. and 2. to the nearest 5%, use the `QiniTable()` you calculated for the propensity and uplift models in Part 1 to calculate the <u>incremental profits</u> you would have obtained in the `expdata_stacked.test` dataset if you had targeted the optimal percentage of customers suggested by each model.

4. Give two reasons for why one model beats the other in incremental profits.