**Case Assignment – Pilgrim Bank A.2**

This assignment is a follow up to Pilgrim Bank A.1. You will use more advanced techniques to build regression models and to make out-of-sample predictions.

\*\* Make sure to depict your results with beautiful visuals. Visreg and ggplot are both great packages for visualization.

Part 1: Interaction modeling

Begin by loading the dataset “pilgrim A2 data part 1.csv”. This is a subset of the pilgrim data which has no missing values.

1. First, run a regression with only main effects testing the effect of demographics and online/offline as predictors of profit. Interpret your coefficients, significance levels and fit statistics. What do the results show?
2. Now use interaction modeling to test whether the effect of online/offline depends on any of the demographics. If so what is the direction of effect and is it a big effect?
3. Choose one interaction or main effect model that you think does the best job of predicting profit. Which variables did you include and why?
4. One of the conclusions from Pilgrim Bank A part 1 was the online banking seems to have a (small) positive effect. Given the new analyses you have conducted, does this conclusion require revision or caveat? Do these analyses suggest any new strategic recommendations for Pilgrim Bank?

Part 2: Experiment Analysis

Begin by loading the dataset “pilgrim A2 data part 2 experiment.csv”. This data shows the results of an experiment that Pilgrim Bank ran to look at the effect of migrating people to the online channel. For each participant in the experiment we observe profit in 1999 as well as in 2000. The column labeled condition reflects the experimental condition, with a 1 reflecting the experimental group and 2 reflecting a control group. 300 Experimental participants were approached and offered an incentive to switch to online. All accepted the offer. 300 Control participants were approached but not given the offer. All remained in the offline channel. (Note this is an idealized scenario for simplicity).

1. Did profit change in the experimental or control group from ’99 to ’00? You should report summary statistics as well as a significance test.
2. Was randomization successful or are there systematic differences between the two groups?
3. What is your conclusion from the experiment? Was migrating people to the online channel effective at increasing profits? By how much? Which test or tests support your conclusions?
4. Note any weaknesses or limitations of the experiment that qualify your conclusions.

Part 3: Out-of-Sample Predictions

\*\*\*Note: For subsequent analyses, use only the experimental group and ignore the control group.

1. To begin, test a few regression models to try to predict profit in ’00 in the experimental group using whatever variables you like. You can test main effects, interactions, nonlinear transformation, whatever you like.
2. Assess fit statistics and other metrics for each model. Which model seems to be doing the best job? What is your evidence for that?
3. Use K-Folds cross-validation to compare some of your candidate models. Did you learn anything from this that you didn’t know from looking at the model parameters and fit statistics? What values did you choose for K (number of folds) and R (number of replications)? Did variation of these values make any difference? (Note: background information about the cvTools package is available on D2L). Keep in mind that cross-validation entails random sampling. If you want to get the same results each time you need to set a seed.
4. Pick your favorite model. **Important: Send your model to Bart and I by Monday night via email. Your email subject line should read: “prediction tournament” Your model should be a single line of r code of the form model1<-lm(…). Your email must also include the names of all group members. Make sure we can easily understand what each variable in your model means. If it is not clear from your variable names, please add a note to your email explaining any variables whose interpretation may be unclear.**
5. Next, load the dataset “pilgrim A2 data part 2 targets.csv”. This dataset shows demographics and profit in ‘99 for 1400 customers -- not currently online -- that Pilgrim Bank is thinking about targeting with a similar intervention to the experimental group in order to migrate them to the online channel.

Use your chosen model to make predictions for the target customers for profit in 00.

You can use a line of code like the following to create a new column in the dataset with predictions:

pilgrim.targets$bestmodel <- predict(model1, pilgrim.targets, type="response")

Should Pilgrim Bank move forward with the intervention with the target customers? Why or why not? How much should they be willing to pay as an incentive to the target customers to switch? Back up your recommendations with analyses.

Finally, note any limitations of your analyses that qualifies your recommendations.