CSC423: DATA ANALYSIS AND REGRESSION / CSC 324: DATA ANALYSIS & STATISTICAL SOFTWARE II

Final Project – Next Steps

Don't wait until week-10 to start on the project – start now, so you have more time to work on it and refine if necessary.

Here some helpful tips –

1. **Domain Knowledge**: Get a better understanding of the dataset, variables and how it relates to your DV. This will not only help with the analysis, but also explain the variables /definitions of the variables for your presentation and in the reports including comping up with interaction variables.

2. **Data Cleaning:** Check the data, clean if necessary and import the data into SAS dataset.

3. **Methodology:** Make a list of the different stages/steps you need to follow to complete the analysis. See the project instructions for more information. This list will come in handy to carryout your analysis and write the methodology section for the final report.

4. **References:** References are useful for a variety of things (See separate document on References_How to cite the.PDF)
   (1) Subject matter – if the references you picked were for similar datasets or industry, you can learn what types of variables they have used and what the outcome was
   (2) Methodology – you can see what methodology they have used and test some of those in your analysis as well
   (3) Compare analysis and results – you can compare their methodology and results with yours to determine which methodology produced the best results
   (4) Future direction – compare your analysis to the references to identify any future direction of the project

   Goal: For the presentation and reports, you will compare and contrast the references with your analysis to see if you saw similar results, similar outcome? If not why, etc.

5. **Checking Performance:** Before you fit the model, split the data into training and testing, fit the model on training set, test the model performance on test set. There will be a 20% reduction in both presentation and final report grade if test set performance is not included.

6. **Responsibility of each team member:** Each person has to do their own coding and come up with their set of model(s) that is distinct from the rest of the teammates. When you split the data into training and testing – make sure to use a different seed value (see notes for info.). Make sure to do this step before you fit the model or carryout the model selection method. This means you will have different models – you may have the same predictors, but the parameter estimates will be different. So you will compare all model performance per step 5, and report it. You may test out additional models as well. Code developed by all members should be submitted for the final deliverable.

7. **Reports (2 reports):** Make sure to divide the work equally. Each person will be responsible for writing the explanation for their model(s). Keep track of what's delegated to each member. You will need to provide explanation in the team evaluation document.