EDA Final Project and Presentation Guidelines

Presentation

Your presentation should be no longer than 12 minutes (might have to modify depending on how many groups we get). You should have a couple of slides prepared to support what you present, and every group member should speak. Since some groups will have to present substantially before the writeup is due and some groups will present just a couple of days before the writeup is due, we will give more leeway to the earlier groups to have work still in progress.

You will be graded on the following criteria:

- Introduction describes the research question well gives motivation for why the analysis was conducted.
- The data is described so that the audience could follow the subsequent findings.
- Models fit or plots presented were described clearly and results were interpreted correctly.
- Conclusion summed up the overall findings of the study.
- Correct length of presentation and polish/professionalism.

Writeups

For the writeup, you should submit two files: (i) a PDF of up to 10 pages, excluding appendices; and (ii) a file with your code. You should also upload your data (talk to me if it's more than 10 MB.) Your graphs should be of readable size when printed. You can use any statistical or graphical technique you want except for p-values. Your write-up will be three-quarters of your final project grade.

You'll be graded on the following criteria (not all of equal value):

- *Statement of goals.* What questions are you trying to address? Why do you care? Why should we care?
- *Description of your data.* In addition to graphical displays, this should include verbal descriptions of what your variables are, who the individuals in your data from, and how they were selected/sampled. If you have many variables, you don't have to describe all of them, just pick out some key ones.
- *Answering your questions.* This is the most important criterion. It will probably include (but is not limited to) fitting a statistical model or models of some kind, and showing that these models tell you something of interest. You should do the following (not necessarily in this order):
  - State answers to your questions;
  - Describe how you came to these answers;

  - Explore the implications to your answers. For example, if your answer is a non-trivial model, plot the fit and describe what's going on in words.

- *Identification of work left to do/limitations.* It's EDA, so we don't require perfection. However, you should have a clear idea of what the imperfections in your work are (what doesn't fit well? what other variables would you really want to know?), and how they could potentially be addressed.
- *Presentation and professionalism.* You should have a readable, informative, comprehensive, clearly labeled set of graphs, and a comprehensible write-up with few glaring spelling and grammatical errors that makes the main points of the analysis clear.
- *Reproducibility.* If we can reproduce your work from your code, you get all the points for this.

Notes and tips

- This is a class project, so there's no need to be too ambitious. Asking a question or two and finding reasonable answers to them should make for an adequate project.
- Your questions don't have to have numerical answers, or even objective answers.
- Don't be afraid to do elementary things. If it's useful to say what the mean is, say what the mean is.
- We're doing exploratory data analysis, which means it's okay to pursue things that turn out to be dead ends. If you hit a wall, it's worth mentioning this in your write-up. Remember, however, your write-up should be much more about your results than about your process.
- Predictive accuracy is not the primary goal here. A model with a small number of explanatory variables may give you better understanding than a larger model that predicts better.