

Instructions for Final Project:

Background and Logistics:

Your final project involves doing data analysis and drawing conclusions (both quantitative as well as qualitative) on a chosen data set. This will be done in your learning teams. (Groups of 3-5 students.) The chosen data set should contain prices, some form of demand response to price (either quantity based, as we've seen in the airline pricing example in class #1, or transaction-based as we've seen in class #3). The data should ideally contain additional explanatory variables, that we referred to in class as "contextual information" that help explain the price variation. The data set should contain "enough" samples to make the analysis possible (let's say at least 1000 data points and ideally more).

Your objectives and analysis steps should include:

- 1) Exploratory data analysis on the chosen data set, as you see fit, and use of regression to provide some insights into the key effects you observe. To wit, what explanatory variables impact the dependent variable, their statistical significance, and the extent to which the dependent variable is sensitive to changes in these variables (including elasticity estimates).
- 2) Formulation of a demand function model which is based on the regression you have run, and optimization of the estimated model to arrive at the estimated optimal static price for your set up.
- 3) Analysis of variation and spread in observed prices relative to what you computed to be the optimal price based on your model.
- 4) Running an online estimation-optimization algorithm (as we discuss in classes #4 and #5 and as explored a bit in your homework) to optimize prices on the fly. Comparing the resulting price path from this algorithm to the observed prices, and the optimal estimated price (from step 2).
- 5) Deriving further qualitative insights that could inform players in the designated market to which the data set pertains and possible use cases.

Your results, insights etc will be summarized and reported in two separate installments:

1. Part 1 will be a presentation in class on our last meeting (Session 6, June 13). **It is mandatory to be present in this session.** The time allocated to this is about 30-40min with some room for questions and discussion. You are limited to no more than 20 slides, and can use the whiteboard or any other medium to communicate your findings. You may split the presentation duties in any way you see fit, though ideally every member of the team should get some air time.
2. Part 2 will be a written report, 10 (single side) pages, typed in 12pt font, single spaced. This includes supporting exhibits (regression results, graphs, diagrams plots etc) and should introduce the problem area, the key results, and the analysis that supports it.

In both cases think about explaining what you did, the logic behind it, what you found, and how this led to your conclusions. Tell an interesting (and statistically significant!) story.

Timeline:

- **Prior to May 27:** consult on nature of data set choice etc.
- **By May 27:** each team needs to submit their final data (by email or dropbox link etc). This should be a cleaned and good to go data set, ready to run a regressions on etc, in CSV or Excel file format, with a short accompanying note explaining what the variables are, what they mean, the time frame and any other information needed to understand the data.
- **June 6:** (a week prior to presentation) In class, brief consultation re nature of slides for presentation.
- **June 13:** slides submitted to me by that date, and presentation due in class.
- **June 28:** final project report due.

All submission of data set (plus accompanying notes), slides, final report can be done via Moodle, email attachment (if size isn't an issue) or dropbox link (etc).

