

MGSC 310

Prof. Jonathan Hersh

Final Project Deadlines and Instructions

General instructions: Your final project should take a real-world data set, and estimate a series of predictive models against this dataset. You should identify a business use-case for the prediction, and estimate at least three predictive models we covered in this class against the dataset.

November 10th – Due: students must upload to Canvas a one-page outline of their project. This outline should include:

- a) identify a dataset you will use
- b) the outcome you are trying to predict, and what variables you will use to predict it
- c) motivation to your project -- as in the business or practical management use case of such a prediction
- d) three methods you will use to analyze your question of interest
- e) the names of the students who will be part of your group
- f) Rmarkdown document showing summary statistics (mean, std dev, min max) against the baseline dataset.

If you have difficulties identifying a dataset or project please reach out to the TA or the professor. If you have a dataset from an internship, consulting opportunity or job you have been meaning to analyze, you are welcome to use this dataset. After reviewing your projects I may suggest alterations to the project.

November 19th – Due: students must upload to Blackboard a compiled Rmarkdown document that shows one model(s) estimated against the data. Code and replication files are required.

Dec 3rd, Dec 8th, Dec 10th Due: a 10-15 minute presentation in class describing and motivating the question to be studied, summary statistics over the data, the analysis, comparison of performance, as well as a conclusion indicating which model should be implemented and why. Students must send the professor and the TA your slides beforehand. I will send a sign-up sheet for your presentation slot later in the semester.

Friday of finals week @ Midnight: Final project code, replication files, dataset, and 4-8 page writeup describing your problem, motivating the problem, summary statistics, analysis, as well as results and conclusion.

Final grades will be assigned based on a combination of accurately applying the skills we've learned in class to creatively address an interesting management problem.

Useful sites to find datasets:

- Kaggle: <https://www.kaggle.com/datasets>
- Kaggle: <https://www.kaggle.com/annavictoria/ml-friendly-public-datasets>
- FiveThirtyEight <https://data.fivethirtyeight.com/>
- TidyTuesday: <https://github.com/rfordatascience/tidytuesday>
- UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets.php>