MSDS 601 FALL 2023
Final Project Guideline

The final project is designed to test the student's ability to perform a complete linear regression/logistic regression analysis and communicate the results to a broader audience. The students are required to find an open dataset on kaggle or other websites that comes with at least 2 numeric variables, 2 categorical variables, and 100 observations. No datasets directly called from Python or R packages are allowed.

Choose one variable as the outcome response, perform a regression analysis (linear or logistic) using the data, and then discuss the results in the final report.

Due Date: Thursday 10/12, 11:59pm

Submission:

- A final report in .pdf, and notebook(s) with all your Python code. The submission will be group submissions
- Peer review from each student.

Report Description:

The final report will be a written report ((in word file,sllides, latex, and etc that is converted into a pdf) that clearly communicates the information below:

- Description of your dataset: resource, dimension, variable description, etc.
- Statement of the research problems, and a summary of methods
- The explanatory analysis may include but not limited to graphs, demographic summaries, crosstables, individual tests, etc.. And some findings at this point.
- Regression analysis.
- Model selection
- The model diagnosis may include but not limited to: assumption validation, influential points, check for heteheteroscedasticity and multicolinearity.
- Make a final model of choice.
- Write a summary of your findings/results, focus on the significance features and assumption check, and how the model results explain your original research question.
- Add a paragraph to discuss the potential problems of the data and results, for example, severe multicolinearity might impact the sensitivity of your tests.
- NO PYTHON CODE IN THE REPORT

Grading rubric:

The project is graded based on a total of 20 points. 16 points are based on the identical report, 4 points are based on the member contribution calculated by peer review: **Each individual student is required to give a peer review to your groupmate.** With a total score of 4, your final grade of this part will be based on the average of the peer scores your teammates give to you. If you give a groupmate with score <4, please include why you think your teammate should not have a score 4.

Report grade (16 total):

- 8 points will be distributed to the required information/methods provided in the report description.
- 5 points will be distributed to the interpretation of output from each method
- **1 point will be given to extra interpretation and creativity**: for example if you have included any method that was not required for the project or extended the project to more research questions.
- 2 points will be given to the summary where you summarize your findings from the case study to a potentially broader audience.