MIS 3335 Data Analysis Using Python
Guide to the Final Exam

The final exam will consist entirely of coding exercises in a Jupyter Notebook. It will be given during the time scheduled by the University. You will open the exam description, download the files, and complete the exam within the two-hour window.

This is an open note exam. When we do an exam like this in the classroom, I allow one page (both sides) of notes to be used during the final. While you will not be limited to one sheet of notes, it is a good idea to approach the final in the same way. More time spent creating a quality note sheet will result in less time and stress on the actual exam.

The final exam will cover the following topics. If you can apply the standards we've discussed in class to complete these tasks, you will do very well on the final. The following topics should be addressed on your note sheet.

1. Create a new notebook file from scratch with our standard structure (including markdown cells). Include a header cell, section headings, and documentation as necessary.
2. Load data from a csv file to a data frame object. Use a value in the data as the index value while addressing dates and potentially missing values.
3. Replace any missing values with one of the following: column mean, column mode, or a specific value (like 0).
4. Change data frame column data types.
5. Use seaborn to create a scatterplot with regression line that includes a plot title, x-axis label, and y-axis label like the example at the right.
6. Boolean subsetting: use a conditional statement on a column to select specific rows of the data and save them in a new data frame. Example: get the rows from the "tips" data where group size is 3 or more.
7. Use GroupBy to create a simple aggregation of a column using count, sum, or mean. Example: find the average total bill for each day.
8. Run a basic multiple regression using statsmodels. Understand if the model needs to be adjusted and run again.
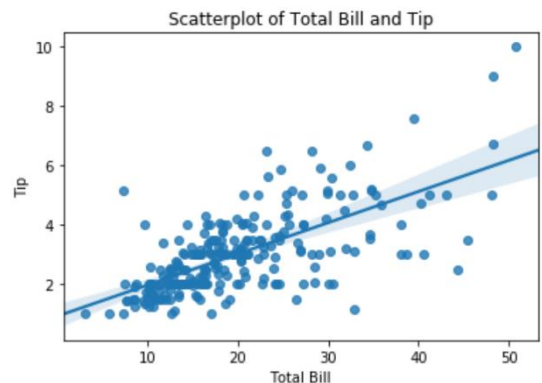
Use methods like head(), tail(), describe(), info(), and others as needed to investigate the data. Follow the course guidelines for writing good code.