This entire assignment is due before 11:59 pm on Thursday, Sept. 15. Submit your work as a **single ipynb file**. While you can discuss homework among yourself, the work must be individual. There must be no sharing of code. In previous years, we have been very strict about plagiarism, and students who were found to have violated the university’s academic policy of conduct were reported.

Before beginning on the homework, make sure to review “HW1 Jupyter basic tutorial.docx” especially the notes at the end. In particular, make sure to name your file as indicated in the notes at the end. We reserve the right to take points off accordingly if the instructions are not adhered to properly.

We will explore a basic issue that arises often in analytics, namely whether to use a one-sided vs. two-sided student’s t-test, and also whether the test should be paired. You may want to review Student’s t-test either from your old statistics text or from Wikipedia (it is accurate in this case). You will be using the data provided in the file we have included with the homework (ISE540\_hw1\_data.csv). Assume that ‘statistical significance’ means a confidence level of 95%.

My primary goal in giving you this exercise is to reinforce the importance of basic statistics and significance testing. Feel free to look up any and every resource on the Web to brush up on what you need to know. We’re aiming for conceptual clarity and application, not memorization.

**1. [10 points]** Read the csv file (ISE540\_hw1\_data.csv) into the [pandas](https://pandas.pydata.org/) data frame. Pandas is an extremely important package that you will encounter time and again if you work with data in Python. This would be a great opportunity to start playing with it!

**2. [15 points]** The Dean wants to know if the average GPA of the students is at least 3.5, the default in previous semesters being that it was below 3.5. Using *g* to represent the sample GPA mean for a semester and *G* to represent the population GPA mean, write down the Dean’s null (H0) and alternate (Ha) hypothesis.

**3. [15 points]** Assuming the population variance of the Fall 2018 and Spring 2019 semesters are equal (but unknown), individually conduct the **unpaired** Student’s t-test for each semester. In either semester can you (individually) reject the null hypothesis?

**4. [30 points]** You are asked to determine if the average GPA has *changed* in 2020 compared to 2019. What do you find? Is the difference between the average GPAs statistically significant? *Hint:* *since there are two semesters in each year, you will have to combine column data accordingly. Pandas should allow you to do this!*

**5. [30 points]** In a conversation with the person who sampled the data, you now discover that the 21 subjects sampled in 2020 (both in fall and spring) are the *same* as the subjects in 2019 (however, you are told that you can still make the equal variance assumption about both semesters). Given this information, you can now conduct a **paired** t-test. Conduct the test, and comment on whether your conclusions from #4 have changed.

***Hint:*** *it should not make a difference for this exercise if the same students are repeated across the fall and spring (of the same year). The only thing that matters is that the students sampled across both semesters in 2019 can be paired with the students sampled in 2020.*

**Useful resource on paired versus unpaired t-test (but there are many others):** <https://www.statology.org/paired-vs-unpaired-t-test/>