STAT 476-576, Spring 2022, Final Project Instructions for Both Undergraduate and Graduate Students

The basic goal for your final project is to analyze some data using Python code. This can be done in one of two ways. First, you could simulate a dataset, which is then analyzed. For example, in the Homework PDF there's a simulation of many games of Chutes and Ladders, which are then analyzed. Second, you can find a dataset and create some appropriate plots, data summaries, and hypothesis tests. For example, see the analysis of Fisher's Iris dataset or the regression analysis of the 'cars' dataset in the lecture slides.

You need to post your proposed final project topic, which must say what dataset you're using (either a simulation or an existing dataset) and an indication of what analyses you plan to do. I'll either accept it as is or make suggestions. Once you get my approval, submit your final project's PowerPoint slides as a PDF file at the end of the semester (see the class web page for due dates.) Make sure your file names are of the form "Final Project YourLastName." For example, I would call my slides "Final Project Bilisoly.pdf" and my code "Final Project Bilisoly.txt." The latter is a .txt file because our email system does **not** allow attaching .py files for security reasons, which can be bypassed by saving your code as text.

Length of Your PDF: If you're taking this class as STAT 476, then please turn in a presentation that's 10 to 15 slides long. If you're signed up for STAT 576, then please make your PDF 18 to 25 slides long.

Sources of data: In the lecture slides PDF, there is an example of using an existing R dataset in Python. Many of these would work for a final project. There are datasets available online, for example, check out Kaggle.com. My advice is to follow your interests and get data related to them. For example, if you like sports, there is plenty of sports-related data online.