Problem set 7

This problem set is an opportunity to get feedback on progress towards your final project. Please use the final\_project\_template.ipynb. You can add cells to this sheet, but please do not delete any cells!

If you are using BeautifulSoup (scrapy) to webscrap data, then please have the following completed:

1. When you save your file, please have WS at the end of the file so that we know in advance that this contains web scraping. For example, Swan\_Garrett\_ps7\_WS.ipynb
2. Write your code for how you will web scrap the data.
3. Show the first 5 and last 5 rows of the dataset
4. Visualize your dataset with at least 3 visualizations
   1. A histogram for the features you are interested in
   2. A scatter plot for any relationships you are interested in (or boxplot if you plan to look at categorical variables)
5. Write a description in a text cell with what you plan to do next. Make sure you address 1) how you will handle feature engineering, 2) what kind of analysis you want to do
6. If you do more than these steps, that is okay! Just note that Adnan will be focusing on grading the steps above this.

If you are importing a csv file or using pandas.read\_html, then please have the following completed:

1. Attach the CSV file to CANVAS
2. Import the CSV file or use read\_html to extract your data
3. Show the first 5 and last 5 rows of the dataset
4. Visualize your dataset with at least 3 visualizations
   1. A histogram for the features you are interested in
   2. A scatter plot for any relationships you are interested in (or boxplot if you plan to look at categorical variables)
5. Start working on your feature engineering
   1. Show whether or not there are missing data
   2. Show whether or not there are duplicates
   3. Show whether or not there are outliers
   4. List the variables you will be dummy coding
6. Write a description in a text cell with what you plan to do next. Make sure you address 1) how you will handle feature engineering, 2) what kind of analysis you want to do
7. If you do more than these steps, that is okay! Just note that Adnan will be focusing on grading the steps above this.