Final Project
Data Science I

For the final project you will be generating three related scientific questions from a dataset of your choosing. You will then use the dataset (along with an additional external dataset) to explore and answer your questions. This analysis will be done completely in python. The final project will be turned in as a formal written report in Jupyter answering the three questions that you posed.

Links to a number of good resources for open-source data can be found in this blog post https://bit.ly/2GEOo1u. But feel free to find data anywhere! You can also use data that you have been working on for your own research.

The steps of the project are outlined below:

1. Find an interesting dataset for the project. Make sure the dataset is rich and complex enough to yield interesting questions.
2. Perform exploratory data analysis with your dataset. All analysis for this report will be performed in python!
3. From the exploratory data analysis develop three questions you will answer from this data
    a. For at least one question bring in an external dataset to merge with your data that will help answer the question.
    b. For one question you must use at least one of the unsupervised learning methods we learned in class to answer the question.
4. Answer the questions. Provide tables and figures that support these answers.
5. Write a formal report of your results as a Jupyter notebook with the code suppressed. The report should be broken into the following sections:
    a. **Abstract** – summarize the entire analysis, with an emphasis on key results.
    b. **Introduction and Background** – provide the necessary background information for the analysis you will perform. Introduce the three questions you will be answering.
    c. **Methods** – describe the relevant parts of the data analysis methodology you employed for this report. How did you clean the data? What python libraries did you use? What analysis did you do in order to answer your questions?
    d. **Results** – What are the answers to the questions. Include the tables and figures in this section and interpret your results.
    e. **Conclusion** – a high level summary of your findings for this project.
6. Include an appendix of your exploratory data analysis in an **Appendix** of the report.

Final project pitch session (11/20/2019): Each person will be assigned to either the instructor or the course TAs for this project. For the final project pitch session, it is expected that the student will have steps 1 to 3 of the project completed and will have a small informal report to review during the session. During the pitch session each student will meet with the instructor or the TA to get feedback and guidance on the first three steps of the project. This will be worth 5 points (around 15%) of the assignment, so please take the pitch session seriously! You should come with a report that has all of your exploratory data analysis and three well outlined questions.

Final product and submission: The code and html output from your Juypter notebook will be submitted via github. The link for the github repository should be turned in on Canvas.

A rubric for the project can be found below:

	Points	**Points Possible**	**Description**
Pitch session (11/20)		5	Small informal report of steps 1 to 3 of the analysis. This includes the exploratory data analysis as well as three well outlined questions generated from the data. Points will be awarded based upon preparedness for the meeting and communication of proposed project.
Exploratory Data Analysis		5	Exploratory analysis of the data. How well is the data visualized in tables, graphs, etc.? How clearly is the data and purpose communicated through these visualizations?
Code		5	How parsimonious, clean, and concise is your code? Is the code commented?
Methodology		10	How well are the questions answered? Was an external dataset used? Was unsupervised learning used?
Written Report		5	Are the required sections included in the report? How well are the methods and results communicated?
Total		**30**	