WEEK #1

CLASS 1: DATA WRANGLING WITH PANDAS

- Pandas for data exploration, analysis, and visualization
- [Split-Apply-Combine] pattern

Homework:

- Pandas practice with [Automobile MPG Data]
- Simple examples of [joins in Pandas]
- Don't forget about the Command line exercises

Optional:

- To learn more Pandas, review this [three-part tutorial] or review these two excellent (but extremely long) notebooks on Pandas: [introduction] and [data wrangling].
- Read [How Software in Half of NYC Cabs Generates $5.2 Million a Year in Extra Tips] for an excellent example of exploratory data analysis.
- Quora: What is data science?
- Data science Venn diagram
- Quora: What is the workflow of a data scientist?
- For a useful look at the different types of data scientists, read [Analyzing the Analyzers] (32 pages).
- For some thoughts on what it's like to be a data scientist, read these short posts from [Win-Vector] and [Datascope Analytics].
- For a fun (yet enlightening) look at the data science workflow, read [What I do when I get a new data set as told through tweets]
- For a more in-depth introduction to data science, browse through these [PowerPoint slides] from Columbia's Data Mining class.

CLASS 2: EXPLORATORY DATA ANALYSIS AND FEATURE ENGINEERING

- Lecture: Visualization
- Lecture: APIs

Homework:

- Exploratory Data Analysis and visualization practice with [IMDB DATA]
- Note: This homework isn't due until Next Saturday.

Optional:

- Watch [Look at Your Data] (18 minutes) for an excellent example of why visualization is useful for understanding your data.

Resources:

- For more on Pandas plotting, read this [notebook] or the [visualization page] from the official Pandas documentation.
- To learn how to customize your plots further, browse through this [notebook on matplotlib] or this [similar notebook].
- To explore different types of visualizations and when to use them, [Choosing a Good Chart] and [The Graphic Continuum] are handy one-page references, or check out the [R Graph Catalog].
- For a more in-depth introduction to visualization, browse through these [PowerPoint slides] from Columbia's Data Mining class.
- [Mashape] and [Apigee] allow you to explore tons of different APIs. Alternatively, a [Python API wrapper] is available for many popular APIs.