Capstone Project Proposal by David Whitney (2/23/2016)

Project Overview: Thousands of American companies makeup American stock exchanges, such as the New York Stock Exchange and NASDAQ. However a recent Wall Street journal article, using data compiled by brokerage firm JonesTrading, stated that actually six companies accounted for more than half of the value added to the NASDAQ in 2015: Amazon, Google, Apple, Facebook, and Netflix, and Gilead [1]. Thus, one reasonable hypothesis might be that the performance of the NASDAQ stock index is actually strongly predicted by the performance of these top stocks, designated in the proposal by the acronym 'FAANGG' (Facebook, Amazon, Apple, Netflix, Gilead, Google, Apple, Facebook). For this capstone project, the student proposes to determine whether the individual or collective performance of these top stocks does a good job of predicting the performance of the NASDAQ. The rationale for embarking on this capstone project is not merely just to address the specific question laid out for this project, but to develop a generalized analysis pathway that could allow the student to arbitrarily compare performance of any selected market index or portfolio (such as the S&P500 or the student's own financial portfolio), and determine what underlying stocks are contributing strongly and weakly to the selected market index or portfolio.

Initial questions to address:

1. Using financial data compiled from Yahoo.com, the student will determine how correlated 'FAANGG' stocks are with the daily performance of the NASDAQ across the last 5-10 years.
2. Although the student will start comparing 'FAANGG' stock prices to the NASDAQ, stock prices are not the only salient metric. The student will also investigate whether other metrics, such as either market value of a 'FAANGG' company or stock volume traded on a given day, are better predictors of NASDAQ performance.
3. Next, the student will address whether linear models combining these 'FAANGG' stocks can robustly predict the performance of the NASDAQ.
4. Utilize more complicated machine learning techniques to determine whether other stocks, not just the 'FAANG' stocks, might be even better predictors of the NASDAQ performance.

Some potential Issues:

1. **Accessing data from multiple sources:** Although the student proposes to initially use the web.DataReader from the pandas python package to import financial data from Yahoo.com, the student may require either writing a customized reader or using another package to access additional financial data. For example, the intrinsic pandas data reader does not appear to provide data for the daily market value of a company.
2. **Incomplete data** – Not all of the 'FAANG' stocks existed for the last 5-10 years. For example, the IPO of Facebook occurred on May 18, 2012.
3. **Data that may need to be modified** – The student will have to determine how factors such as stock price splits, or the addition/subtract of stocks to the NASDAQ might impact the robustness of the student's analysis.
4. **Collinear data** – Stock prices for a given company are not likely to be independent from the stock price of other companies, and thus the student will have to consider carefully the impact of collinearity.

[1] D. Strumpf. Wall Street Journal. "The Only Six Stocks That Matter" http://www.wsj.com/articles/the-only-six-stocks-that-matter-1437942926 (July 26, 2015)