**Read Me**

**Table of Contents:**

* **Project Description**
* **Data Collection Scripts**
  + Opensecrets APIs
  + Opensecrets scraper
  + NY Times election data scraper
  + Quandl API
* **Data Files**
  + **Political Contributions by Industry** 
    - Opensecrets API data
    - Opensecrets scraped data
  + **Stock data**
    - Dow Jones Industrial Average daily data
    - S&P 500 daily data
    - Sector-Specific Indices
  + **Election results**
    - FEC election results
    - NY Times 2014 election results
* **Data Dictionaries**
  + **Stock Ticker Symbol Lists**
    - S&P 500 stock tickers
    - Dow Jones Industrial Average stock tickers
  + **FEC election result prefaces**
* **Cleanliness Assessment**
  + **Cleanliness Assessment Script**
    - Script analyzed
  + **Cleanliness Score Results**
    - **Factor one:** Outliers, per any numbered values 3 standard deviations from the mean were counted as outliers and the percentage was calculated by feature. The outliers percentages were averaged for the data source score
    - **Factor two:** Missing Data
      * **2.**1. if there was no data in a column cell it was counted
      * **2.**2 for strings, if the number of capital letters didn't match the numbers words that was also counted and percentage was calculated combined with empty cells
    - The Missing data percentages were averaged for the data source score
  + Both Factors were averaged and then subtracted from 1. The higher the score, the cleaner the data

**List of Files:**

* **Project Description.docx (not on github yet):** This document describes the data science problem, analyses that might be conducted using the merged dataset, and issues with the data.
* **Opensecrets API.py:** This script obtains data on political contributions by industry for 2012 and 2014, for general election winners, from Opensecrets.org, using the getLegislators and candIndustry APIs.
* **ScrapeOpenSecrets.py:** This script scrapes data on political contributions by industry for 2004-2010 general election winners and losers and for 2012-2014 general election losers, from Opensecrets.org.
* **GetNYTimes2014ElectionResults.py:** This script scrapes 2014 general election results from the NY Times website.
* **Project 1 Financial Historical Data.py:** This script obtains S&P 500 and DJIA daily data from 2004-2014 from Quandl.com
* **Contributions by Industry 2012-2014.csv:** 2012 and 2014 political contributions by industry. Output of Opensecrets API.py.
* **FundingCongress.csv:** Political contributions by industry for 2004-2010 general election winners and losers and for 2012-2014 general election losers. Output of ScrapeOpenSecrets.py
* **FinancialData.zip (not on github yet):** Daily data from DJIA, S&P 500, and sector-specific indices, 2004-2014. Outputs of Project 1 Financial Historical Data.py, in addition to separately downloaded sector-specific index data.
* **FEC Election Results.zip:** FEC election results from 2004-2012 primary and general elections.
* **NYTimes2014elections.csv:** 2014 general election results. Output of GetNYTimes2014ELectionResults.py.
* **nytimesTemps.csv =** due to format issues, it had to be converted in order to better process through the cleanliness script.
* **SP500.csv:** Stock tickers for components of S&P 500.
* **dowjonesIA.csv:** Stock tickers for components of DJIA.
* **FEC Data Dictionaries.zip:** Prefaces to FEC election results.
* **Clean Analysis Script.py:** Assess cleanliness of results files.
* **Cleaning Analysis output.csv:** Cleanliness scores for all datasets. Output of Clean Analysis Script.py