Step 1. Normalize all of the data (N-O/O)

Step 2. Issue: Markets open and close, means some stock data has a value of 0.0 (meaning no change). Resolved issue by ignoring all data fields were the stock data field shows a 0.0, worked.

Step 3. Calculate the correlations for the various lags, just iterate through the data and take

Stock\_data\_i with polarity\_data\_(i-lag), worked perfectly, save all of the lagged correlations (set the amount of lag to 12 maximum which is 1 hour of lag).

Step 4. Got all of the lagged correlations (and unlagged) in separate files, combine them together by taking the average for every *lag* over all the files present. (for example: take lag 0 in EVERY FILE, then iterate through all of the files and take the average correlation coefficient. Also calculated the respective standard deviation). Did this for my own classifier and the textblob classifier. Also did the same exact thing for the number of tweets to see whether there exists a correlation between those

Step 5. Additional regressions? Additional data?

Questions:

* Is taking the average of all of the correlations a good measure for the *actual* average correlation? + What exactly does the standard deviation tell in this case? The percentage change it can be above or below the average as calculated?
* Regressions to run besides this? Other ways to form the data?
* Ways to display the data? (example present in the Excel file)
* Maybe try implementing a trading portfolio based on this? What rules should be used (when to long when to short)?

**Own classifier:**

Total tweets counted: 368956

Total tweets counted on average per day: 6708.3

Greatest Negative Polarity Change: -88.3575883576% 2017-04-21 20:21:13.412453 20:26:13.412453

Greatest Negative Stock Change: -0.880553588341% 2017-03-27 15:25:30.449058 - 15:30:30.449058

Greatest Positive Polarity Change: 642.857142857% 2017-05-09 21:25:13.23280 - 21:30:13.232802

Greatest Positive Stock Change: 0.432159061597% 2017-03-10 15:27:45.525880 - 15:33:45.525880

**Textblob classifier:**

Total Tweets Counted: 2211716

Total Tweets Counted on Average per Day:40213.01

Greatest Negative Polarity Change: -84.6666666667% 2017-05-19 15:34:44.774808 - 15:39:44.774808

Greatest Negative Stock Change: -0.880553588341% 2017-03-27 15:25:30.449058 - 15:30:30.449058

Greatest Positive Polarity Change: 500.0% 2017-05-19 15:29:44.768276 - 15:34:44.768276

Greatest Positive Stock Change: 0.432159061597% 2017-03-10 15:27:45.525880 - 15:33:45.525880

Overall tweets analyzed: **2580672**