a) Objective:
The purpose of this project is to get a better understanding of stock volatility in the healthcare industry. Simply by making observations, you can see how erratic the behavior prices can be. That in and of itself was the initial reason to pursue this study. Then there is of course the advent of new technologies that make for an interesting moment in the history of market behavior. Lastly, it is not enough to do statistics on this kind of data. There are plenty of available tools in the python language waiting to be utilized for the purpose of forecasting prices. The ultimate purpose is to create a simple methodology in preparation for a predictive model with sole desire to accurately predict prices in the healthcare industry.
b) Data Summary:
Retrieval of stock data couldn't be any easier. The yahoo api makes the process very smooth. All you need to do is chain a pandas data reader to the yahoo data method and you'll get a multi indexed data frame with as many columns of company stocks as you wish. You will need to retrieve tickers on your own. You will want to take the freshest of tickers from the yahoo finance webpage. The only trouble is that if you wish to analyze hundreds of tickers, you will encounter read errors. Some of the tickers will present an error so try except is necessary. That aside, the process of getting tickers into a data set over a specified time period is simple and manageable.
c) Methodology:
The issue of selecting stocks from a large pool is still a subject of debate. Portfolio analysis does Markowitz optimization on a set of stocks with weights. These weights represent a percentage of your budget invested in each of your stocks so that the sum of the weights is 1. This approach says nothing about selection prior to weight optimization. As it stands, it is not so easy to find literature on this subject but you might be able to guess that the best means of approaching the subject would be through some kind of simplified statistical analysis. The methodology taken can be reduced to a few steps. The first step is checking a stock returns normality through kurtosis and skew. The second is to calculate sortino ratio. This is not a necessary alternative but it is an alternative to the sharpe ratio. These ratios are tradeoff measures between risk and reward. The chosen alternative does not penalize positive swings in return behavior. Simply choosing the most positive or least negative (for those returns that churn out negative only ratios) of sortino ratios. Choose however many you wish to study or model and proceed from there. That concludes selection. Modeling would be the last phase. Since the time series approach was chosen initially, investigating correlation and stationarity is crucial. One does this with available methods such as the PACF or ACF. You may even wish to use the Dickey Fuller as we did, to check for stationarity. Once you have it has been confirmed that the time series is stationary, we

can move on to the modeling phase. Autoregressive methods were used such as TBATS, and AutoArima. We even included the LSTM, a recurrent neural network.

d) Possible Takeaways:

Since the healthcare industry will be undergoing changes with the advent of new technologies both statistical as well as biomedical, it is a great idea to begin analyzing the market in preparation for those changes. Building new models appropriate to this industry will give clients new insight into the stability of the industry as well as how to encourage its growth.