Section 4: Conclusions

1. **Describe how your model responds to the problem and include your rationale for conclusions (make sure your data supports your conclusions).**

   Conclusions / Limitations:
   From the sentiment analysis we saw that the majority of the tweets from this period were negative. This is not terribly surprising since most individuals only tweet when there is an issue. Given that this period is also very busy for open enrollment signups there may be more people than normal experiencing issues.
   The correlation analysis had unexpected results - a negative correlation between the number of negative tweets and daily change in stock price. These results must be interpreted with extreme caution though due to the very small number of dates for which tweets were available and the presence of an outlier in the data. It does appear with more data a pattern might be uncovered.

   Some Limitations:
   - ✓ Twitter's API only allows for searches to return the previous 7 days' worth of tweets. This severely limited the number of days available for analysis. In the future this data could be collected longitudinally and stored until sufficient data was available to redo analysis.
   - ✓ Blue Cross Blue Shield is a generic term for insurance offered through companies who have an agreement with the blue cross blue shield association. Consumers often use terms like "BCBS", "Blue Cross Blue Shield" or any number of Anthem's state level plan names (e.g. "Blue Cross Blue Shield of Georgia") to talk about Anthem on social media. Unfortunately, these terms are too general to include in this analysis because they are not clearly talking about Anthem. * "Anthem" is a very general term (as opposed to Walmart or iPhone) so searching for #Anthem returned many tweets that had nothing to do with Anthem the company. I excluded #Anthem for this purpose.
   - ✓ The sentiment analysis performed here was very simple. The very presence of a negative or positive word does not mean the tweet itself is positive or negative (i.e. sarcasm).
   - ✓ The correlation analysis here used logistic regression. Given that stock prices are essentially time series data this type of analysis might not be the best way to correlate stock prices and sentiment scores. While not sure of the exact method that might be more appropriate - non-parametric (signed rank) test come to mind as a possibility.

2. **Describe how you model and findings could be useful to management strategy in the area you selected (this may confirm or be different from expectations set in Part 1).**

   Stock Market prediction is the art of estimating the future value of companies' trade or other financial instrument that traded on exchange. More specifically, stock markets movements are analyzed or predicted in order to guide investors the best time to buy or sell the trade.

   Stock market plays a remarkable part in the economy of a country so it is very vital to identify the factors which has an effect on stock market trends. A talk about a product or a business news have a greater influence on the markets, but it is really unknown how accurately it affects the market movements. This project visualizes the relation of news sentiments and stock market trends. R language's available functions and algorithms helps to facilitate rapid data analysis and visualization and it very easy to interface with new data sources and other computing environment in the financial industry. There are more than 70 packages in R that are presumably developed directly solve the problems in computational finance.

   The primary function of stock market is to support the growth of industry and commerce in the country. Rising stock market is a sign of developing industrial sector and a growing economy of a

country It will also affect the pension funds because most of the private pension funds will invest in the stock market. Movement in the stock market can have an economic impact on everyday people. A collapse in the share prices have widespread economic depression. Stock market makes the stock as a liquid asset unlike real estate investment.

Fundamental Analysis involves the in-depth analysis of companies past performance, it involves understanding the company apparently in terms of its product sales, man power quality, infrastructure, profitability on investment. It also considers revenues, earnings, future growth, return on equity, profit margins, social acceptance and other data which determines the companies underlying values and potential growth for future. Technical Analysis is the science or skill of forecasting the future financial stock trends based on an evaluation of past price movements. Like nature events forecasting, technical analysis does not end in hundred percent prediction result about the future. Instead, technical analysis may help traders to think what is exactly going to happen to prices over time.

Sentiment analysis can work but should be used in conjunction with fundamental analysis. Often you can get a feel for turning points based off of sentiment but it's often after the fact. For example, if there's a lot of Angry News about a prototype for a new iPhone, it could influence a change in the stock price (if the info wasn't already priced in) so you want to take it with a grain of salt. In an efficient market sentiment analysis would not be a sound strategy because everyone would be following it and the info would already be accounted for. In an inefficient market, if you could reasonably figure out Twitter sentiment that would lead to a price change you could make some gains if your information is substantial.

3. **Describe what next steps and what else needs to be done to improve your model and findings.**
   Finally, its worth mentioning that this analysis doesn't take into account many factors. Firstly, the dataset doesn't really map the real public sentiment, it only considers the twitter using, English speaking people. It's possible to obtain a higher correlation if the actual mood is studied. It may be hypothesized that people's mood indeed affect their investment decisions, hence the correlation. But in that case, there's no direct correlation between the people who invest in stocks and who use twitter more frequently, though there certainly is an indirect correlation - investment decisions of people may be affected by the moods of people around them, i.e., the general public sentiment. While training the sentiment analyzer, 770 tweets are used which is comparatively a less number to train a sentiment analyzer. In future, we look forward to human annotate more than 10,000 tweets and train the classifiers. With increasing size of training datasets, the models tend to perform better. All these remain as areas of future research.