Can Twitter sentiments predict stock prices?

Sentiment Analysis

Who? What? How?

Customer/Client: Corporate Public Relations department

Problem Statement:

Gauge how change in company executive leadership is received by general public and investors by looking at twitter and in turn affect company's stock prices.

How would this help the clients/customers?

Client will be able to determine if the decision to change the top leadership is perceived as a good step. In addition to gauging general negative, neutral or positive sentiment, determine how company's stock prices would be impacted in the near future.

Data gathering, preparation and Analysis steps:

1. Tweets from 2 days before the announcement of CEO leadership to 5-6 days after the announcements are pulled from data sources.
2. The data is stored in MongoDB to be reviewed and updated with relevant flag and topics manually.
3. Relevant tweets from MongoDB are retrieved and wranged for exploratory data analysis.
4. Generate Sentiment score via Vader and TextBlob.
5. Apply weight to the sentiment scores by multiplying by how many times the tweet was favorited, quoted, replied to and retweeted.
6. Resample by bootstrap method to increase the size of the sample to 500.
7. Simple OLS linear regression with sentiment scores as predictor variables to gauge stock price trend. Stock price at the close is used.
8. Perform Bayesian Linear regression to predict sentiment scores for future days.
9. Compare the trend for sentiment scores against actual stock prices after the leadership change announcement.

Data set:

Dataset	**URL**	
Twitter - Full Archive search	https://developer.twitter.com/en/docs/tweets/search/overview/enterprise	Tweets from up to 2006. 100 results per search. ● WFC - Tweets from 09/25 - 10/02 ● SAP - Tweets from 10/09 - 10/16
Stock Price - Yahoo Finance	https://finance.yahoo.com/quote/	Exported the history data to csv file

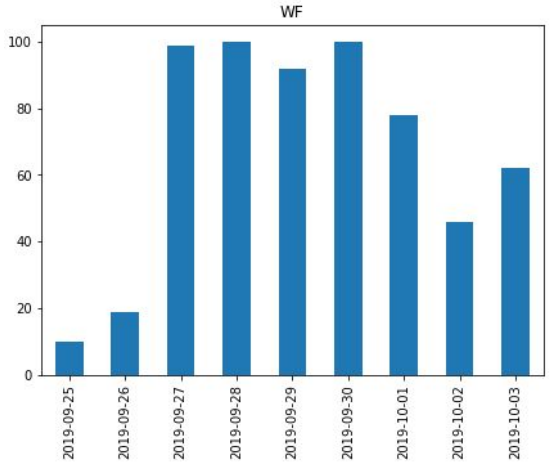
Limitations encountered during data gathering and analysis:

1. Limited number of examples of leadership changes. Only two companies, Wells Fargo and SAP were used.
2. I was only able to get 100 tweets per search at a time using historical search.
   If there were more than 100 tweets a day matching the search criteria, search brings back the most recent tweets. I am also limited to how many free searches I can make with the free account I have with Tweet API.
   Due to this, I retrieved the last 100 tweets of the day for each company ending at mid-night of the day. Depending on how many tweets met the criteria, 100 tweets for the day can be for the whole day range or within the last 20 minutes of the day.
3. Even though search criteria was refined to get the most relevant tweets as possible, before the announcement of the leadership change, most tweets were not relevant, not even with the company itself, thus the number of tweets available for analysis was extremely small for number of days.
   In many cases, tweet content only comprises of link to an article without any other content making it impossible to know what the tweet was about thus contributing to less number of relevant articles.
   Following the link and analyzing the content of the article is not in scope of this project.
4. Tweet contents tend to be sarcastic or use jargons and in-jokes that sentiment analysis package were often not able to score correctly.
   For example, “NBA-style free agency comes to banking", “Wells Fargo Selects Charles Scharf To Make History As The First Female CEO Of Major U.S. Bank”.
   It would have helped if there were functionality to customize the sentiment analysis package by adding custom list of words.
5. There can be more than one issue that can affect the price of the stock (and the sentiment for/against a company) at a time. In this project, only one issue - change in CEO leadership is acknowledged. However, this one issue might not be enough to accurately and entirely describe the stock trend.
   For example, quarterly earning report was announced at the same time the CEO leadership change. Quarterly earnings report was exceptional which might explain why the stock price trend was positive while sentiment scores for SAP CEO trended slightly negatively.

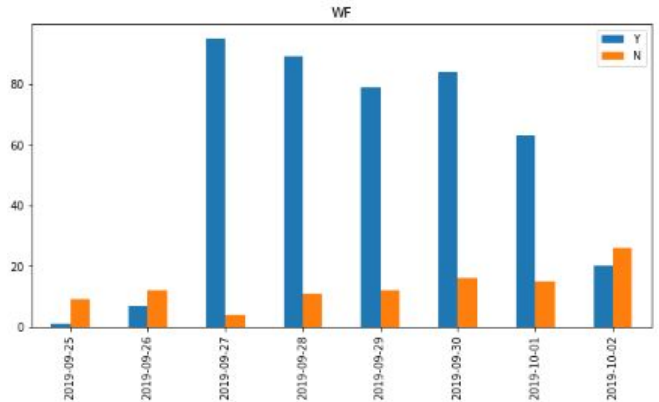
Wells Fargo - New CEO announced on 09/27/19.

Twitter activity is up sharply starting 09/27, the day of the announcement, and so does news articles. While tweets continue to be more than before the announcement, new articles activity falls almost to the level before.
In case of tweets, number of there is sharp increase in relevant tweets starting from 9/27 and the number of relevant tweets don't go below the number of not-relevant tweets until 10/2.
This shows that announcement of new CEO did not go unnoticed and public is interested for at least 5 days after the announcement.

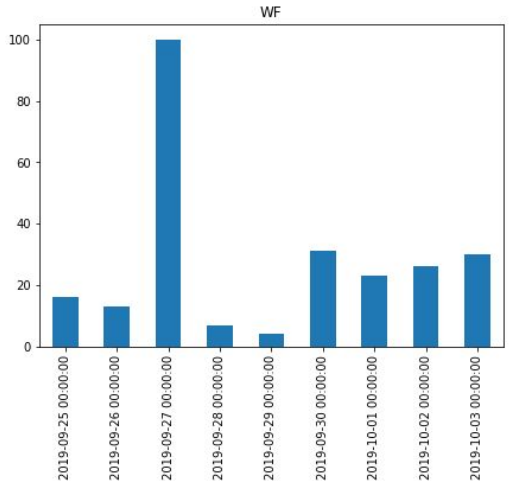
The top topics are Dimon, Diversity, New York, Pay, Stock and Tough Job. This tells me the story that people focused on Charles Scharf having been Dimon's former mentee, yet another white male CEO based in New York. Stock went up modestly, people acknowledge that this will be a tough job given the circumstances and the reason why Scharf accepted the position is because of significant pay bump as an enticement.



# of Tweets per day.



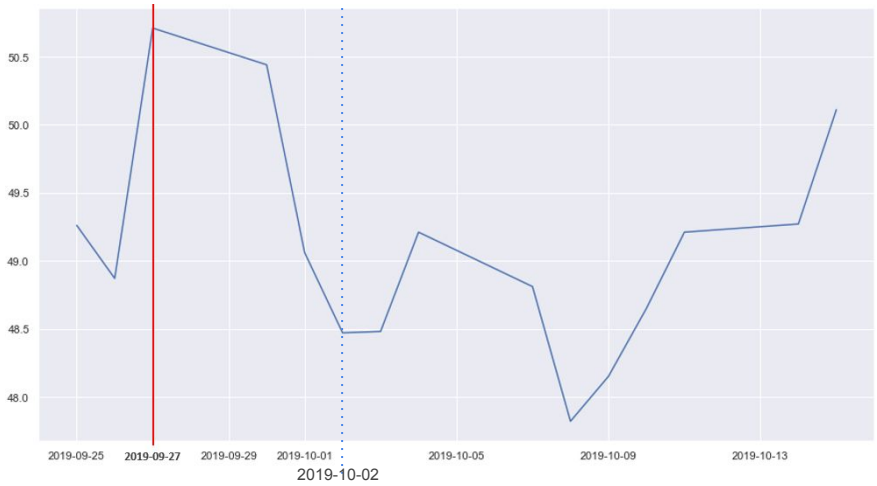
# of tweets per day - divided by relevant (Y) and not relevant(N)

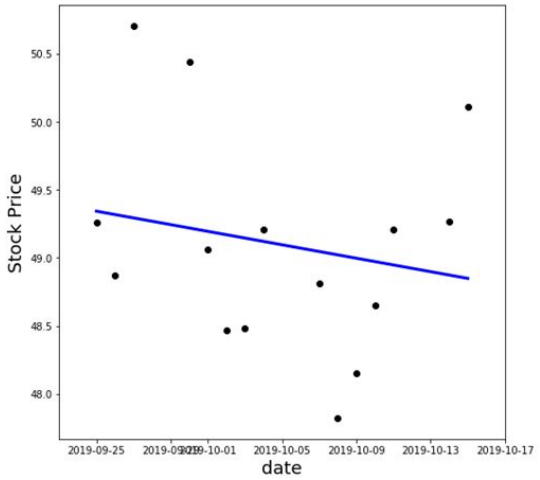


# of News articles per day

Wells Fargo - Stock

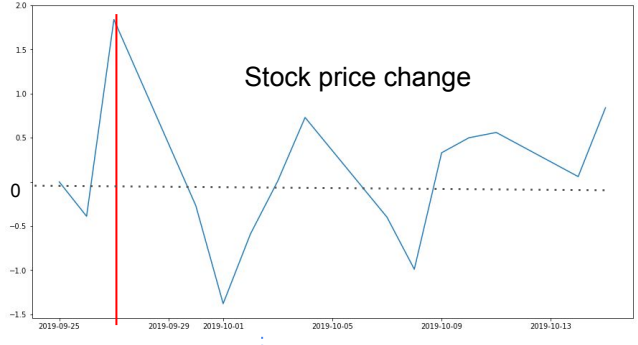
Stock prices jumped by close of 09/27 after the announcement was made but the rally was short-lived and stock went down as much as 5-6% by 10/8. The stock price went back up again after 10/8 but that could be due to a different issue.
Looking at 09/25 - 10/15/19, there is a slight downward trend (linear regression slope -0.02).





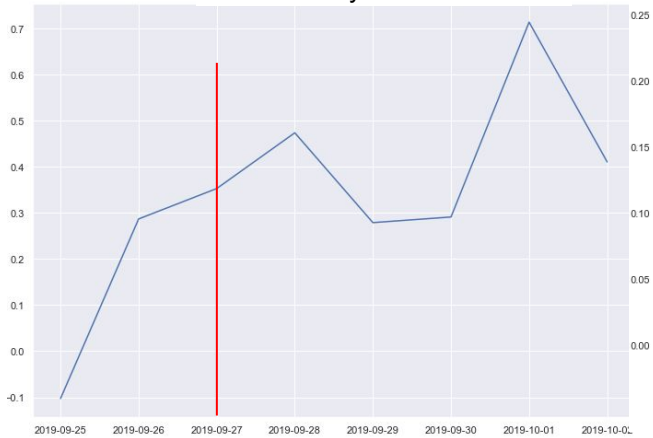
Wells Fargo - Twitter sentiment

Sentiment scores are widely distributed within each day (09/25/19 - 10/2/19), both for Vader and Textblob sentiment scores but Vader score has slightly upward trend (linear regression slope: 0.06) while Textblob Polarity score has slightly downward trend (linear regression slope -0.01).

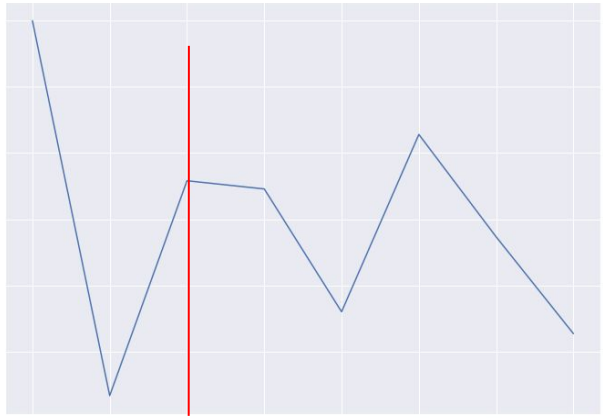


Vader sentiment score trend doesn't match the stock price trend.
Both sentiments went up in score towards positive score on the day of 9/27 which matches the jump in stock price.

Vader - daily mean



TextBlob - daily mean

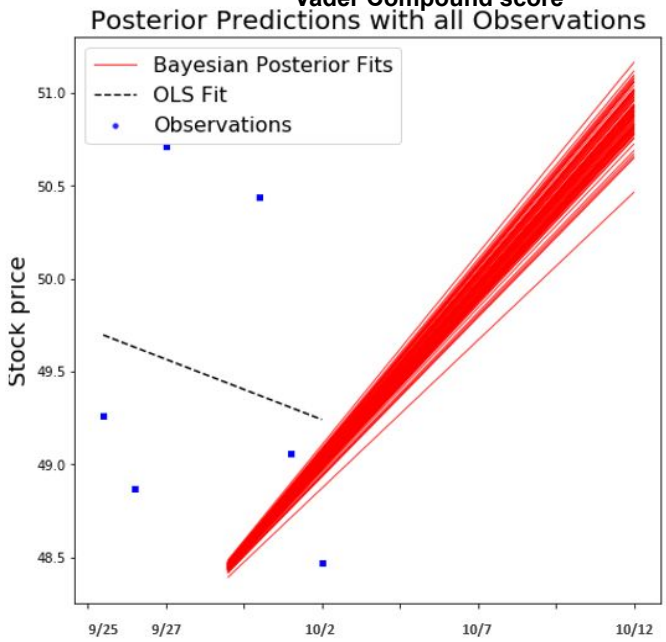


Wells Fargo - Bayesian Linear regression & prediction

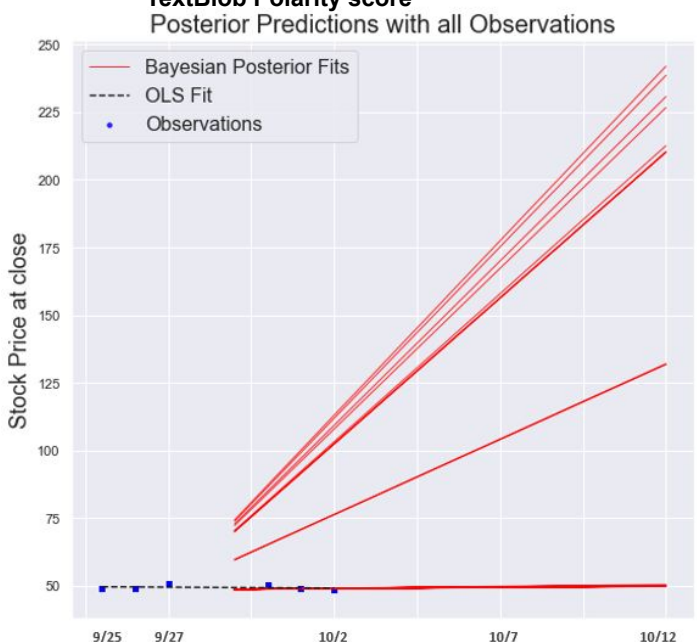
OLS Linear regression predicts downward trend for stock price, however Bayesian linear regression predicts upward trend for the stock price after 10/2.

Actual stock price at close for 10/3 was 48.48. (10/2 was 48.47). For 10/3 stock price at close, with Vader compound score, 49.18 was predicted while Textblob Polarity predicted around 50. 10/2 close price was 48.47 while 10/3 close was 48.48 so the trend was predicted correctly.

Vader Compound score



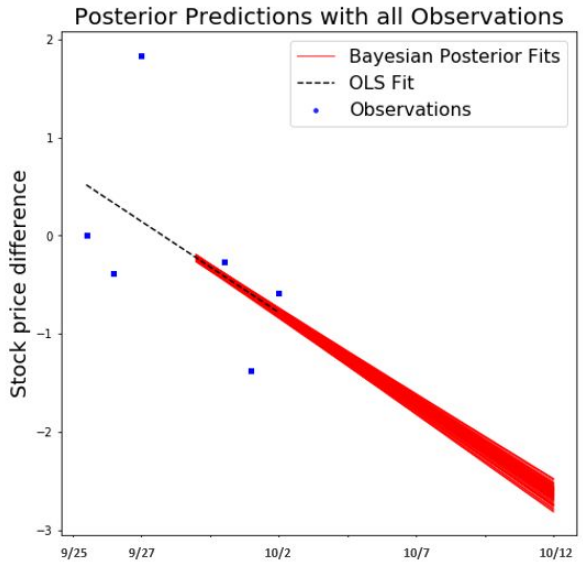
TextBlob Polarity score



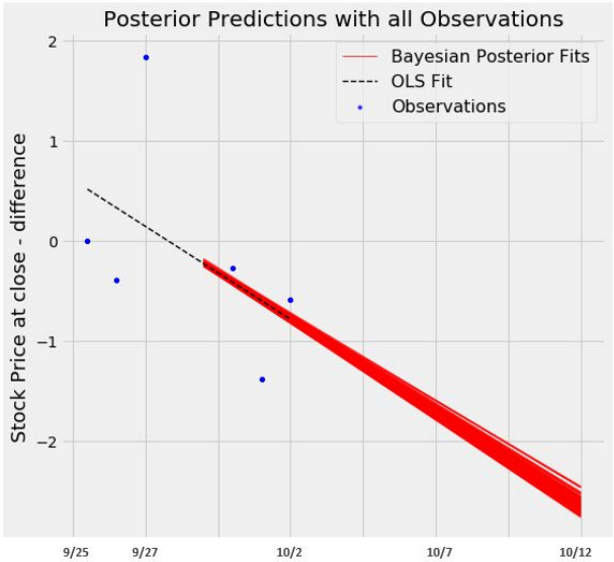
Wells Fargo - Bayesian Linear Regression & Prediction - Stock price change

Same step was taken to see how changes in stock price from each day impacted the prediction. SAP's stock price changes fluctuated throughout the period with a slight downward trend. Bayesian linear regression mirrored this trend which is different from how it behaved when actual stock price was used.

Vader Compound score



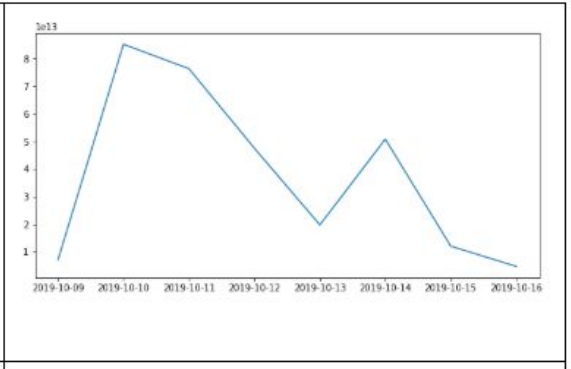
TextBlob Polarity score



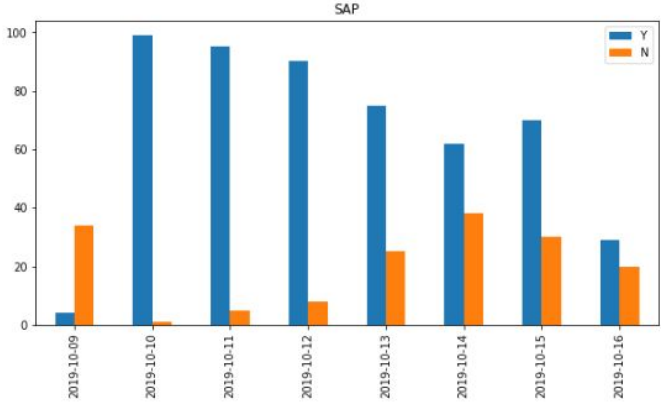
SAP - New CEO announced on 10/10/19 US time (10/11/19 in Germany).

SAP announced the change in leadership on 10/11 in Europe which was 10/10 in U.S.
Tweets started increasing on 10/10 and due to the large amount of activity, there were more than 100 tweets for the day. For the days where there were more than 100 tweets, 1st chart shows the range of time collected for the day (inverted. The amount of time 100 tweets were generated was less on 10/10, the day of the announcement, and started falling in the next few days and down close to the amount of activity before the announcement on the 6th day after the announcement.
Similar to Wells Fargo, the number of relevant tweets compared to not-relevant tweets sharply grow the day of the announcement and slowly dies off towards the 6th day of the announcement.

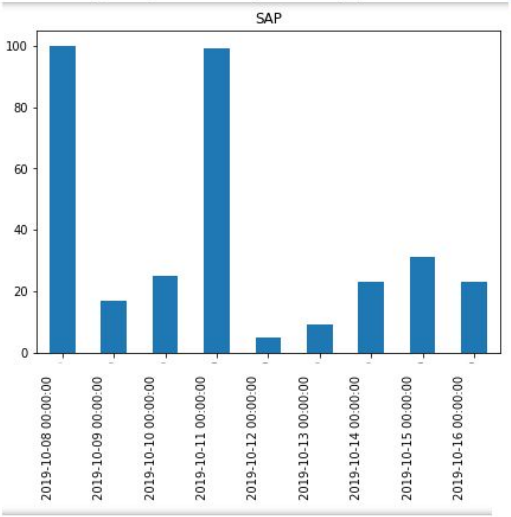
Top topics for SAP are Diversity, New CEOs/New Era and Previous CEO. People focused on previous CEO's achievement and new co-CEOs. One of the co-CEO, Jennifer Morgan is the first woman to lead DAX index company.



1 day - range of time 100 tweets were collected for the day. (inverted)



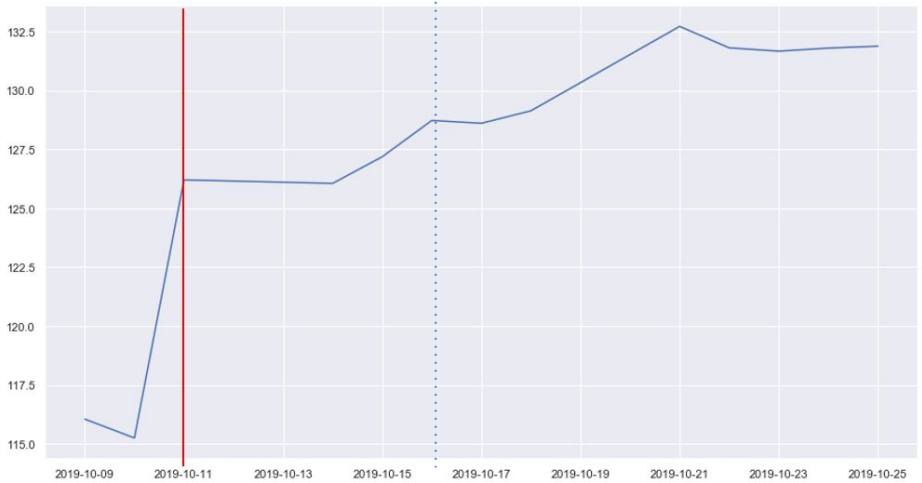
# of tweets per day - divided by relevant (Y) and not relevant(N)

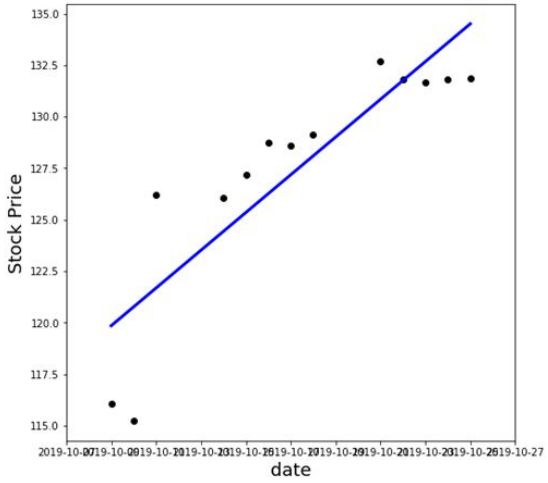


# of news articles per day.

SAP - Stock

Stock prices jumped by the close of 10/11 after the announcement was made on 10/10 US time and the stock price stays to trend positive (linear regression slope: 0.91). Even after the 10/11, the slope was 0.51.
One thing to note is that a very strong earnings report was announced at the same time as the CEO announced his immediate resignation so there is multiple factors impacting the stock price and I think the impact of CEO resignation may not be as as strong.

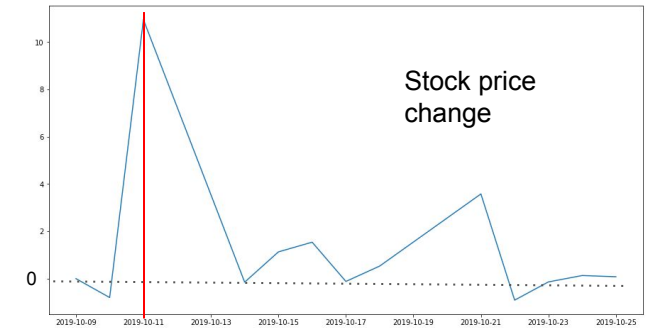




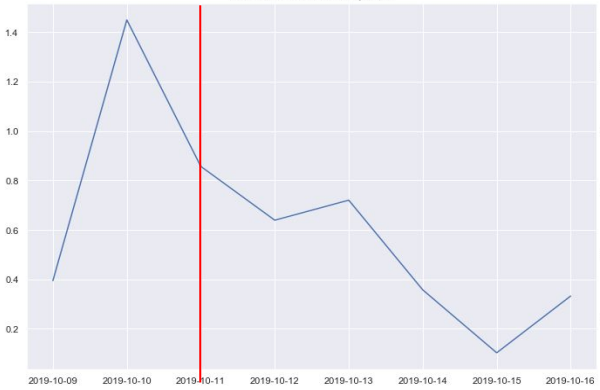
SAP - Twitter sentiment

Sentiment scores are widely distributed within each day (10/09/19 - 10/16/19), both for Vader and Textblob sentiment scores and they both were slightly downward trend (linear regression slope: -0.10 for Vader and -0.025 for TextBlob.

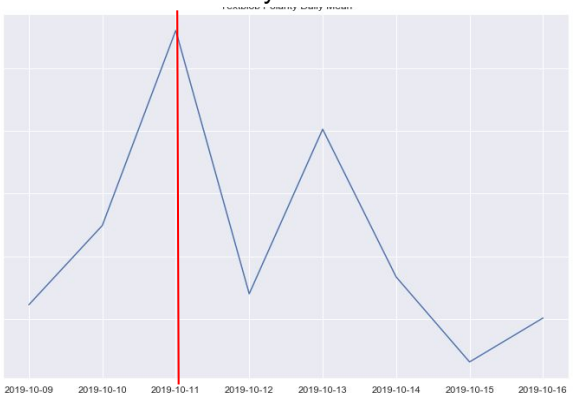
Both jumped towards positive sentiment on 10/10 and 10/11 respectively, none of the sentiment scores seem to be in line with stock price trend at least for the few days after the announcement. As mentioned before sentiment scores may trend downward because the CEO stepped down unexpectedly and the current CEO was considered quite effective and well respected. Still, the downward trend was very slight, considered neutral.
This is why I think the impact of the CEO resignation is not as strong as that of stellar earnings report to the stock price.



Vader - daily mean



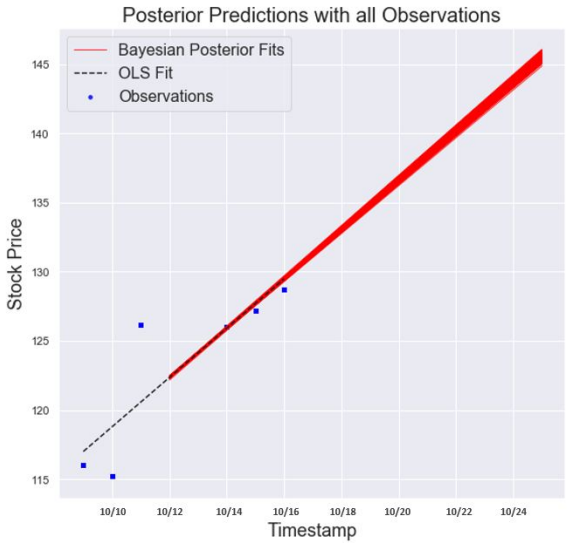
TextBlob - daily mean



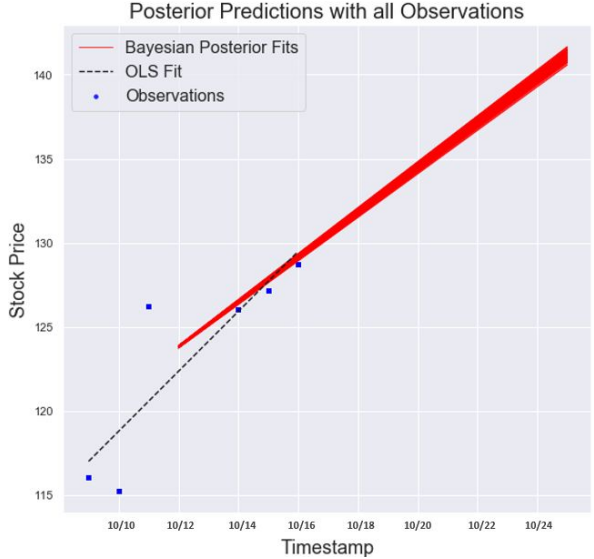
SAP - Bayesian Linear regression & Prediction

OLS Linear regression predicts upward trend for the stock price, and Bayesian linear regression also predicted upward trend for the stock price after 10/16.
For 10/17 stock price at close, with Vader compound score as predictor variable, 131.26 was predicted while Textblob Polarity predicted around 130. 10/17 close price was 128.6. while 10/16 close was 128.72 so predicted price was higher but was not too different. Overall trend for the stock was up and the predicted value fits the overall trend.

Vader Compound score as predictor variable



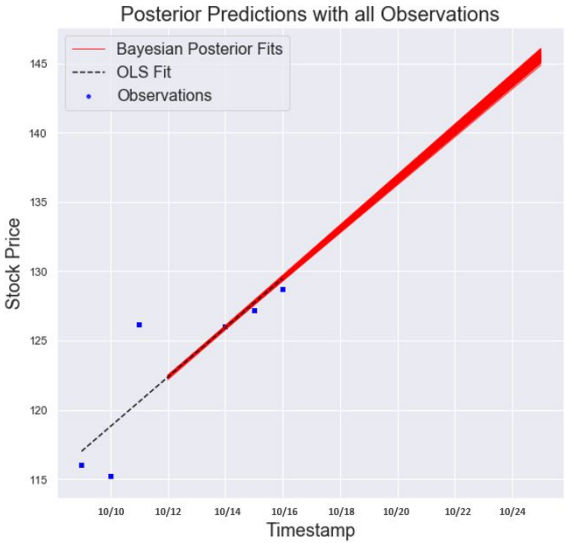
TextBlob Polarity score as predictor variable



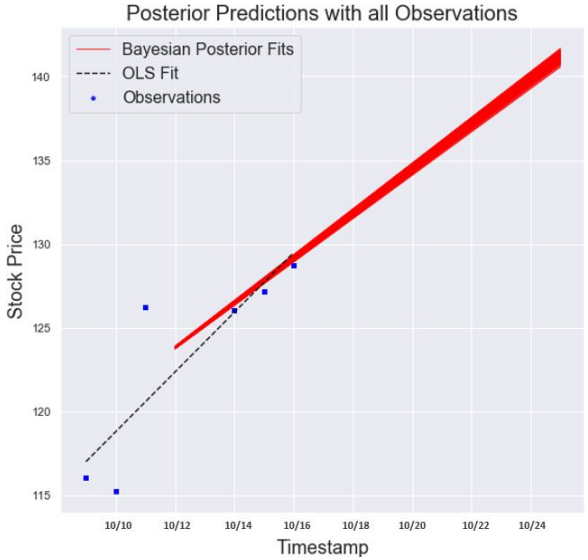
SAP - Bayesian Linear Regression & Prediction - Stock price change

Same step was taken to see how changes in stock price from each day impacted the prediction. SAP's stock price increase slowed after 10/11 and Bayesian linear regression mirrored this trend.

Vader Compound score



TextBlob Polarity score



Conclusion

Sentiment scores and the stock prices jump on the day or the day after the announcement but the impact of the leadership change doesn’t seem to last more than a couple of day or so for stock prices and attention of the general public.

Additionally, there are other factor that impact stock prices as we noticed with SAP where the CEO resignation happened at the same time as the earnings report.

Given these, it is reasonable to expect public sentiment gathered from Twitter is a guideline but not a definite factor in determining near future stock price, except for a stock price within a day or two of the announcement.