The Impact of Social Media Presence on the Stock Price of The Coca-Cola Company

Narpal Sandhu, Srujay Korlakunta, Josh Yeung, Kevin Pham

December 1st, 2017

I. Introduction

Our aim was to better understand how the social media advertising and promotions based on their following rate influence stock markets growth and scalability. We will be examining the stock market through Coca Cola stock (KO) which is a under the New York Stock Exchange. We chose to evaluate the stock of Coca Cola because in the past few years they have acquired multiple companies and have been incorporating their marketing strategy to boost the newly acquired company's sales. The major trends that we noticed is that Coca Cola has reached out to major social media influencers on Instagram, Twitter, Facebook, and YouTube, sponsoring them in order to influence that guide the millennial generation. Since social media information is consumed at such a rapid pace we wanted to see if these marketing tactics are a good use of resources. We defined a successful tactic as one that increases the income of a company and furthermore increases the stock proportional to their social media following and marketing money spent.

Through this project we allow others to apply this method to other companies--to measure how their use of or lack of social media affects their stock prices, and thus their company income. In addition, answers can be found about the effectiveness of specific social media platforms for a company, related to their stock prices.

II. Related Work

1. Influence of Social Media over Stock Market by Juan Pineiro Chousa

Piñeiro-Chousa, J., Vizcaíno-González, M. and Pérez-Pico, A. M. (2017), Influence of Social Media over the Stock Market. Psychol. Mark., 34: 101–108. doi:10.1002/mar.20976

This research analyzes investors' activity through social media and these media's influence over the Chicago Board Options Exchange Market Volatility Index (VIX) using a logit model and a fuzzy-set qualitative comparative analysis (fsQCA). The logit results show that social media sentiment influences stock markets. Meanwhile, the fsQCA results show that the investors' profile is important for explaining how social media influence the stock market. Particularly, holding period combined with experience in technical investors contributes to avoiding a raise in market risk, whereas for nontechnical investors message sentiment and experience form the combination that contributes to avoid a raise in market risk. The difference between this and our project is that while this focuses on when an investor should buy and sell based on sentiment analysis of a stock in his/her own portfolio, our project is geared toward whether or not social media itself can impact a stock price.

2. Sentiment Analysis on Social Media for Stock Movement Prediction by Thien Hai Nguyen, Kiyoaki Shirai, Julien Velcin

Nguyen, Thien & Shirai, Kiyoaki & Velcin, Julien. (2015). Sentiment Analysis on Social Media for Stock Movement Prediction. Expert Systems with Applications. 42. . 10.1016/j.eswa.2015.07.052.

This paper presents the novel method to integrate the sentiments in social media for the prediction of stock price movement. This research proposed a method using the sentiment of the topic for stock market prediction, and is the first research to show the effectiveness of incorporation of the sentiment analysis by investigation on a large scale test data. From a practical point of view, although the average accuracy is only 54.41%, the proposed method can predict the stock price movement with more than 60% accuracy for a few stocks, and performs much better than other methods for the stocks that are difficult to predict with only past prices. In the end, however, this paper only measures whether stock is up or down, and doesn't present a holistic visualization of the overall change in stock prices along with social media use over long periods, which is the main focus of our investigation.

III. Data

This investigation required us to collect 4 major data sets that would be explored in relation to each other:

1) Coca Cola (KO) Stock Price, 2008 - 2017
2) Coca Cola Twitter Follower Count, 2008 - 2017
3) Coca Cola Instagram Follower Count, 2008 - 2017
4) Coca Cola YouTube Follower Count, 2008 - 2017

Follower count data for Twitter, Instagram, and YouTube was manually acquired from WayBack Machine, since the respective APIs do not provide historical follower counts. In WayBack Machine, follower counts were taken from 2 snapshots of the site for each month that there was a record, and averaged to make a standard for the month. For this section, cleaning the data required running linear regressions to approximate follower count growth for the months that did not have snapshots on the site. This allowed us to have a complete, graphable follower count growth graph for each of the social media platforms.

The raw data for the KO stock was taken from the Yahoo Finance API, which comprised of 3 columns: date, number of shares, and closing price. To clean the data, we extracted the closing price for the first day of each month, to match the number of data points in the follower count graph. Also, since stocks of major companies normally spit-off (for example, a stock that is worth $50 dollars can be broken down to two $25 dollar stocks to increase the amount of people that are able to invest and if the company is growing), we had to ensure that the API was able to level the playing field by averaging the stock price based on the total number of shares available. The Yahoo Finance API was used in Python, and we wrote a method that would extract the stock price for each consecutive day in a time frame given the stock symbol, start date, and end date. This was done using the datascience and pandas libraries in Python as well.

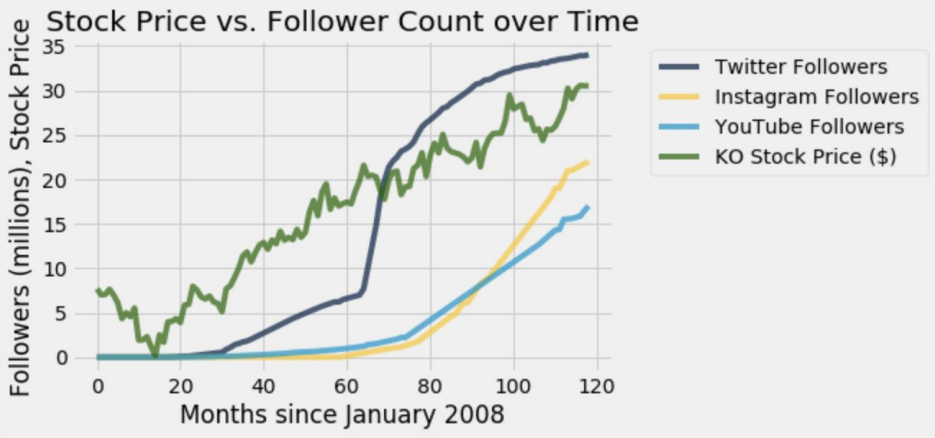
The result of all this data collection was a table that had a data point for each month from January 2008 to November 2017, with that month's Twitter, Instagram, and YouTube follower counts, as well as that month's stock price (**cocacoladata.csv)**. This data was also then used to calculate the growth rate of each data set for that time frame.

References:
Stock data taken from Yahoo Finance API: https://pypi.python.org/pypi/yahoo-finance
Follower Data taken from WayBack Machine http://web.archive.org

IV. Visualization

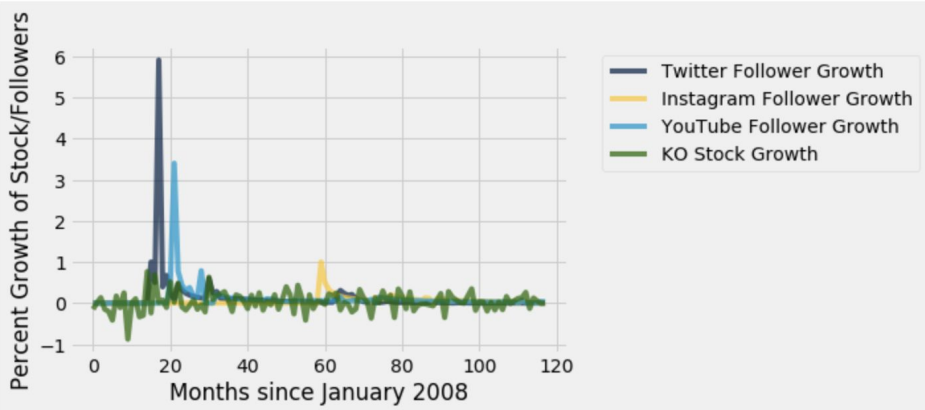
There were 2 main visualizations in this investigation that allowed for a thorough exploration of the topic. The following visualizations were constructed using using matplotlib in Python on a Jupyter Notebook.

1. **Graph of KO Stock Price vs. Coca Cola Follower Counts**

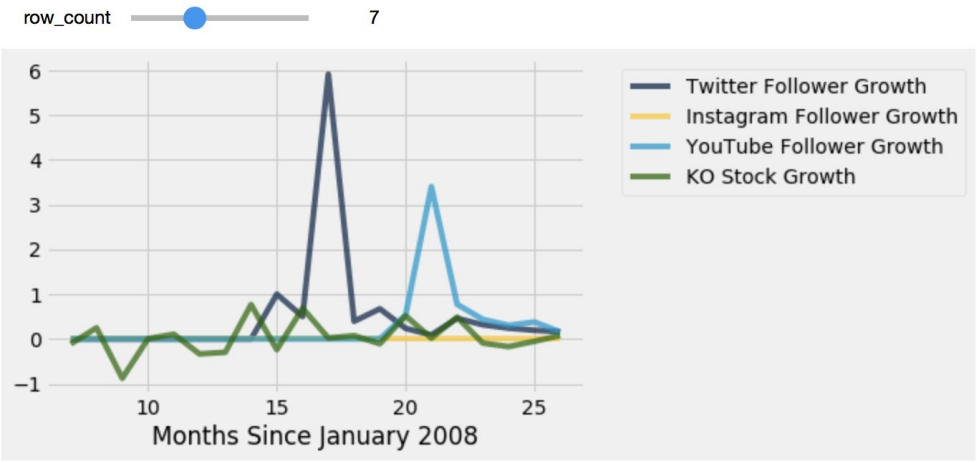


This graph was made by plotting the data for stock price from January 2008 to November 2017 alongside the follower counts of Coca Cola on Twitter, Instagram, and YouTube. This graph shows that there is a strong association between stock price and follower count in the time frame.

2. **Graph of Growth Rates of KO Stock Price and Coca Cola Follower Counts**

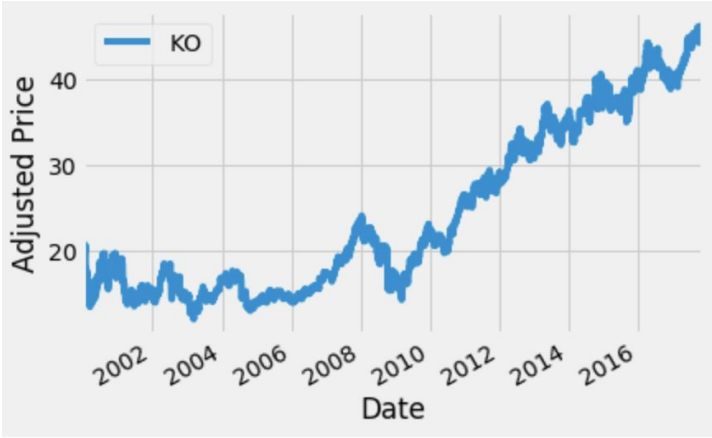


This graph shows the growth rates of KO Stock vs. follower growth on Twitter, Instagram, and YouTube between January 2008 and November 2017. Although it may be somewhat hard to see, it shows that whenever there is an increase in the growth rate of the follower count, there is a relatively similar growth rate in the stock price, even though it may not be as pronounced. Because it is difficult to see details from this graph, we also constructed an interactive visualization that can be seen below. The slider on the top allows you to see a more detailed view a 20 month span, starting with the month that is row_count months after January 2008.

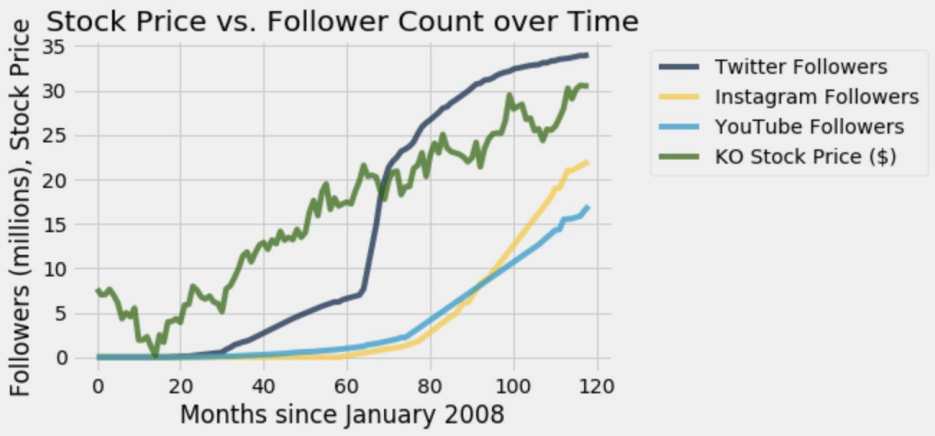


V. Conclusion

In this investigation, we aimed to determine if there is an association between the follower count on social media of a company (specifically Coca Cola), and their stock price. The visualizations that we produced provided us with valuable information that we can use to draw conclusions from the data.



The graph above is a graph of the stock price of KO from 2000 to 2017, and it shows that until around 2010, the stock price of KO hovered around $15 to $20, and in the 10 years before had very little drastic changes except for a slight spike in 2009. The real growth came after, in 2010, when the stock price grew steadily for the last 7 years with slight fluctuations.



The graph above is our first visualization, and it shows that round 25-30 months after January 2008, which would be somewhere around the beginning of 2010, was when the social media follower count of Coca Cola on Twitter, Instagram, and YouTube started growing, along with the stock. This leads us to believe that this increased online presence, with an increased focus on marketing to a new youth demographic, led to the increase in the stock price of the Coca Cola Company (KO).

VI. Final Thoughts

This project gave us a lot of experience in programming/data analysis in a practical setting, where there is no starter code and the data is not all perfect and pre-cleaned. One of the main problems that we faced was having to figure out a way to get historical follower counts when they were not provided. There were a few different solutions we tried, including scraping data from a few websites, but most of the methods did not provide us with enough data to make a solid conclusion. In the end, we had to be creative with our solution, and although it wasn't perfect, we were able to get the data we needed and make a conclusion. We learned that there may not be a straightforward or easy way to get all the data you want in the form you want, so you have to try different things and find a solution to your problem, and the solution may not be perfect. In this scenario, although the method to extract stock data can be used for any company, the follower count extraction must be done manually each time. However, we felt that this practical experience was very valuable to our future projects working with data.

VII. References

Our methods were Python and statistical concepts learned from this course and from Data 8, as well as the professor's presentation regarding WayBackMachine. Data was taken from the Yahoo Finance API, as well as WayBack Machine.