Process Book - Studying Industry Stocks with Company Events

Updated: May 8th, 2017

Brad Davis, Keegan Howard, Cole Retzloff

Overview

The purpose of our project is to get an insight into how certain dates, events, and times affect the aggregate movement of stocks. Our data visualization project consists of real time data and historical data starting early, 2010. Comparing technology companies, i.e. Apple, Microsoft, and Google, we are using our dashboard to give real updates and connections between how these companies benchmark against each other and how market events affect the market stock price of each. We originally looked at a free data source dating from 1998, however, we want our dashboard to be tied to real time information, which this source did not have. Overall, our goal is to create a tool to find insights into how people decide to trade. The stock market is usually seen as a random investing strategy, but we hope to add a little prediction power based on how the market trended in the past.

Our current project is looking at only the tech industry with Apple, Google, and Microsoft as the three stocks being compared. We have improved our analysis power by using a sentiment analysis of twitter feed, utilizing the NASDAQ data to benchmark the position of each tech company, and inputting product release and quarterly earnings to see how these affect each company respectively.

Related Work

We decided to look at historical Yahoo finance data as well as add in real time data because forecasting the stock market is very unpredictable. Historical events are able to predict to a certain extent, so by discovering underlying trends based on financial release dates or the introduction of a product, we can track trends and have a better idea of when to invest. We have narrowed our scope to only one large industry with a selection of three companies within the industry. We all have a shared interest in the stock market, as well as some investments in these

specific companies, so this visualization could prove to be helpful and informative as we move on post graduation. From the start we had a very large idea of what we wanted to, especially focused on the stock market. We wanted visuals that compiled everything into one table that was customizable per person for date range, large events, specific company, and even possibly had options to overlay averages of past stocks. We found an idea of how to work on this from Scottrade, here.

Questions

How does the announcement of financial data or occurrence of large events affect the stock market within these companies? Do they affect each company differently or similarly?

How do certain products of companies affect the stock price? Does an announcement raise/lower? Quarterly releases affect price/ average volume? How does each stock compare to the NASDAQ market change?

How well does Sentiment Analysis predict trends in the overall industry? Is it a reliable indicator of financial performance? Does it predict the overall market, or just companies referenced in sentiment?

Initial Idea of Project & Project Development

We started this project with the overall goal of having a new perspective on stock market trends. We wanted to see if different dates in the news or with the companies affected how the stock moved that day, week, month, and even year. We wanted to use Tableau as the support tool to explain our historical stocks and have the ability to visualize current feature of both the market and the specific stock while also being able to have a little predictive power. Our initial designs incorporated standard features like highs and lows of stocks, historical pricing, 52-week low/high, volume comparative to other companies, daily changes in prices, and year over year change by quarter or year. We then wanted to add a feature that could show significant events

like a crash in the market, release dates of company financials, or significant news about the industry or company. Our initial designs are below:

Figure 1 - Positive/Negative Candlestick Graph

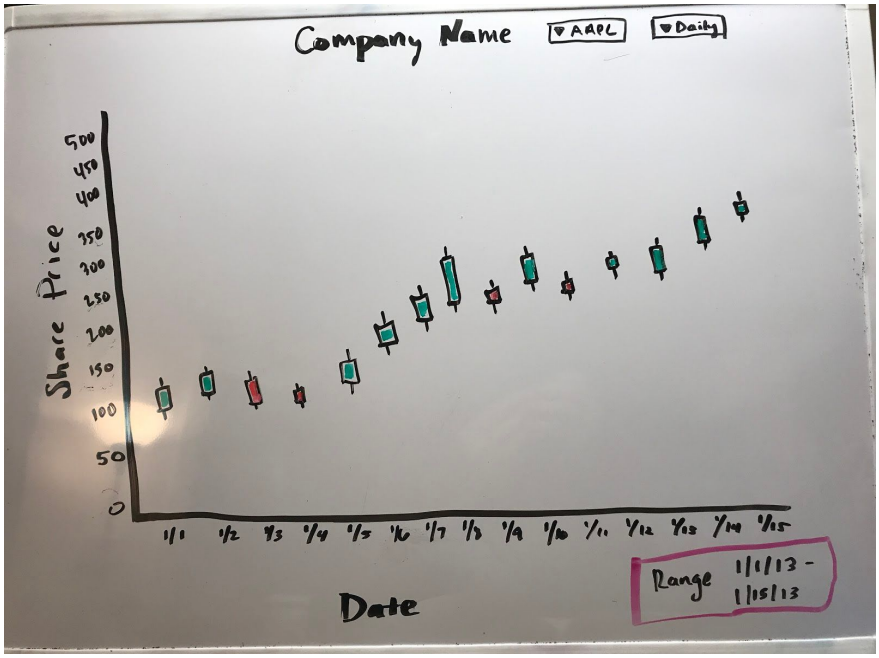


Figure 2 - Date slicers with changes and historical stock data

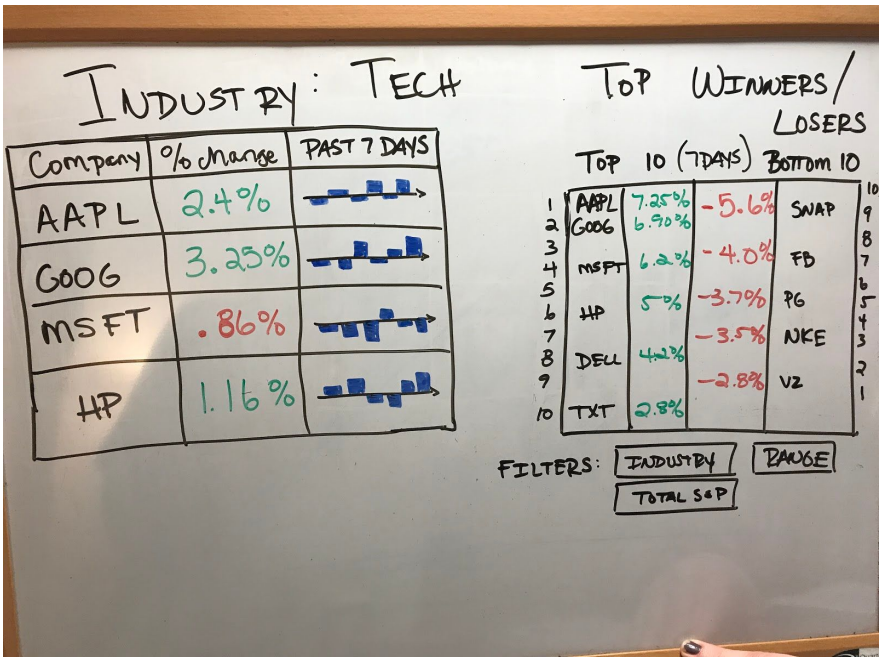


Figure 3 - Comparable stocks in the industry with a possible line graph with events

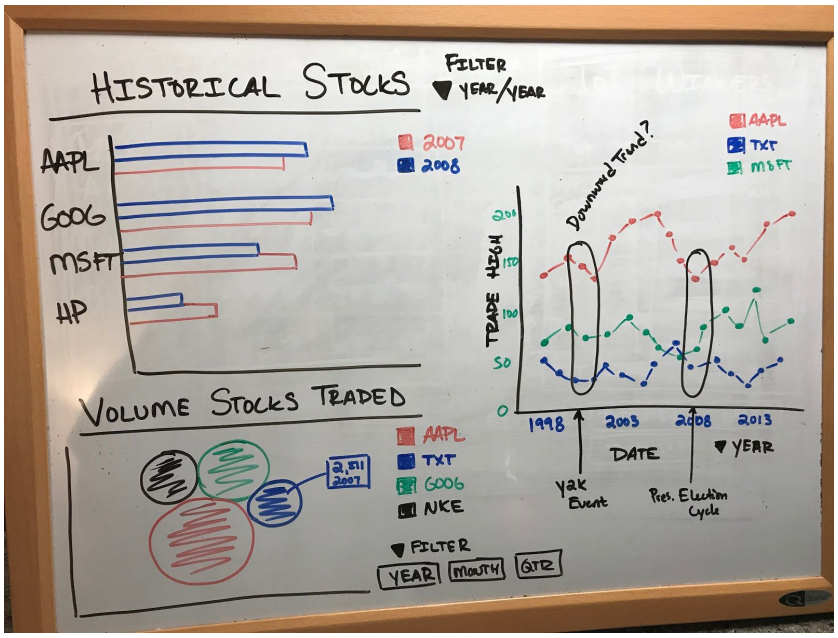
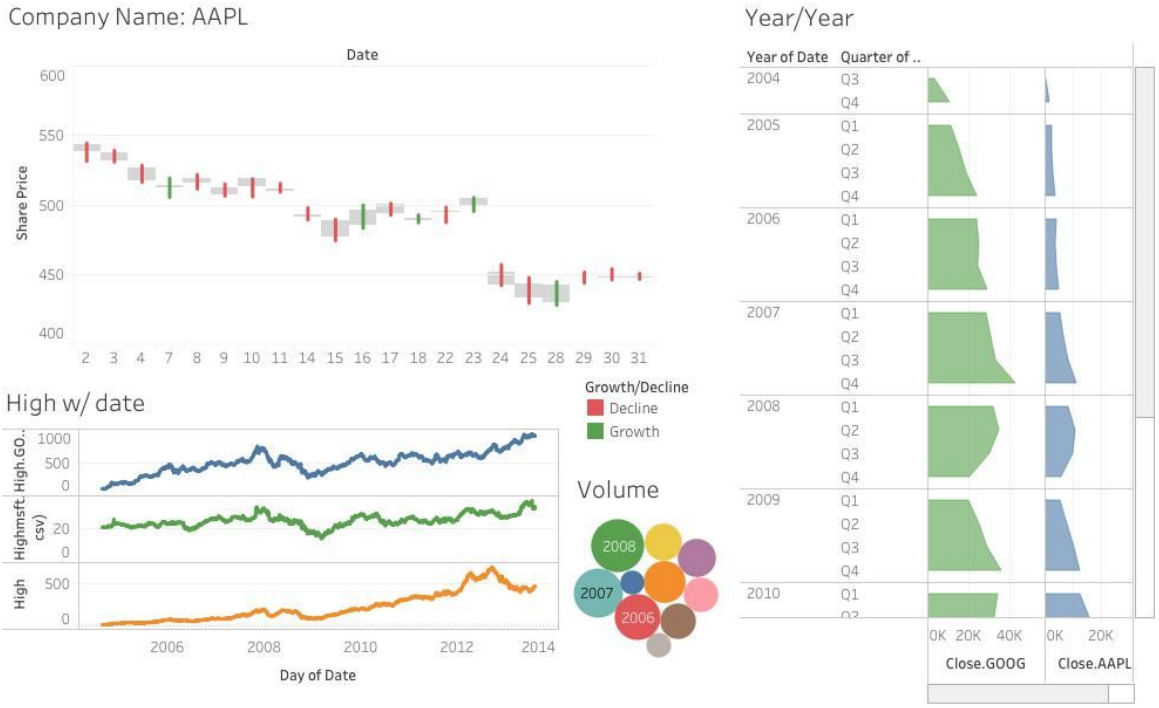


Figure 4 - Tableau example of beginning dashboard



Data Retrieval

Originally, we downloaded and worked with a historical dataset from QuantQuote that began in 1998 and ended in 2013. However, after working with Tableau for some time, we were stuck and wanted a more dynamic tool. After learning basic Power BI in class, we realized how easy it is to connect multiple datasets within Power BI. We looked for new datasets, and after looking into some historical data, we decided to pull real time data from Yahoo Finance instead.

We struggled for a while to pull any live data into Power BI. We are now proud to say that all data in our dashboard pulls is live except for earning releases because there were no online sources for that data that would pull into Power BI. Oftentimes data that we wanted live would pull in as web element that wouldn't work. Thankfully we found out that any tables from websites automatically pull and live update when you request it. This helped out with pulling in Wikipedia data on Apple product launches and will also make sense in the future for pulling info on Microsoft and Google data points.

Upon realizing that Power BI's query editor system is very powerful and useful, we ran with it. Defining steps to data over and over again every time you update from the source is key for updating data for our project. Yahoo Finance's main historical price data page only loads 100 rows at a time which proved to be a challenge until we found a blog that generates a Yahoo Finance link that auto downloads a CSV with the data that you want along with whatever attributes you want. For this solution, we selected historical data from January 1, 2010 to April 10, 2017 along with the data attributes included in the 100 most recent row data pull. Certain letters and numbers in the URL indicated data that you could pull or not pull. For example "j1" added to the URL indicated that you wanted a column with market capitalization (shares x share price). This saved a lot of data cleaning steps because we could add or delete whatever we wanted from the URL. We used this web source to add historical data to the first link which always pulls 100 of the most recent trading days. Upon combining them we would have the most up to date info merged with the historical static source from which we removed duplicates.

We also used another web source called Info that pulls in key indicators from the customizable Yahoo Finance link that we didn't want associated with dates because the values only change every once in awhile including examples such as earnings per share estimates for next year or 52 week high among others. Another Yahoo Finance web source pulls in the stock symbol and last trade price and change in price in order to provide real-time data for our stock ticker on the bottom of the visualization. There are several other minor sources of data including a table with a downloaded image of each company's logo along with their symbol. This symbol picture is used in our slicer to replace the boring text that we originally had. Additionally, another table pulls info from Wikipedia on new Apple product releases which can populate over the adjusted closing price chart on our dashboard. These product releases along with earnings release info can both be overlaid onto the adjusted closing price graph and are shown with red line markers on the graph.

Our final huge data source pulls live tweets from Twitter that include the text $AAPL, $GOOG, or $MSFT. This was no easy task because Twitter seems to be very protective of their data. The only way that we could pull live tweets was to use a website called Zapier. This website detects from an account you link any new tweets matching your criteria and then posts to your online Excel account a new row with whatever attributes you want about the tweet. We decided that important attributes about a tweet include the text itself, the number of retweets, the number of followers which judges a person's influence, their user bio, and the number of favorites. While we didn't include all of these in our project 1 submission, we may be able to utilize more attributes to determine how influential a tweet is. With the text of a tweet, we ran this data through Microsoft's online machine learning sentiment-scoring application that gives a sentiment score between 0 and 1, with 1 being positive and 0 being negative. We believe that this groundbreaking tool will give us an idea about how people are feeling about a certain stock in real time and hopefully predict movements in the price.

We imported several custom visualization templates that were all tested and were judged if they were better or equivalent to the standard ones in Power BI and were found on

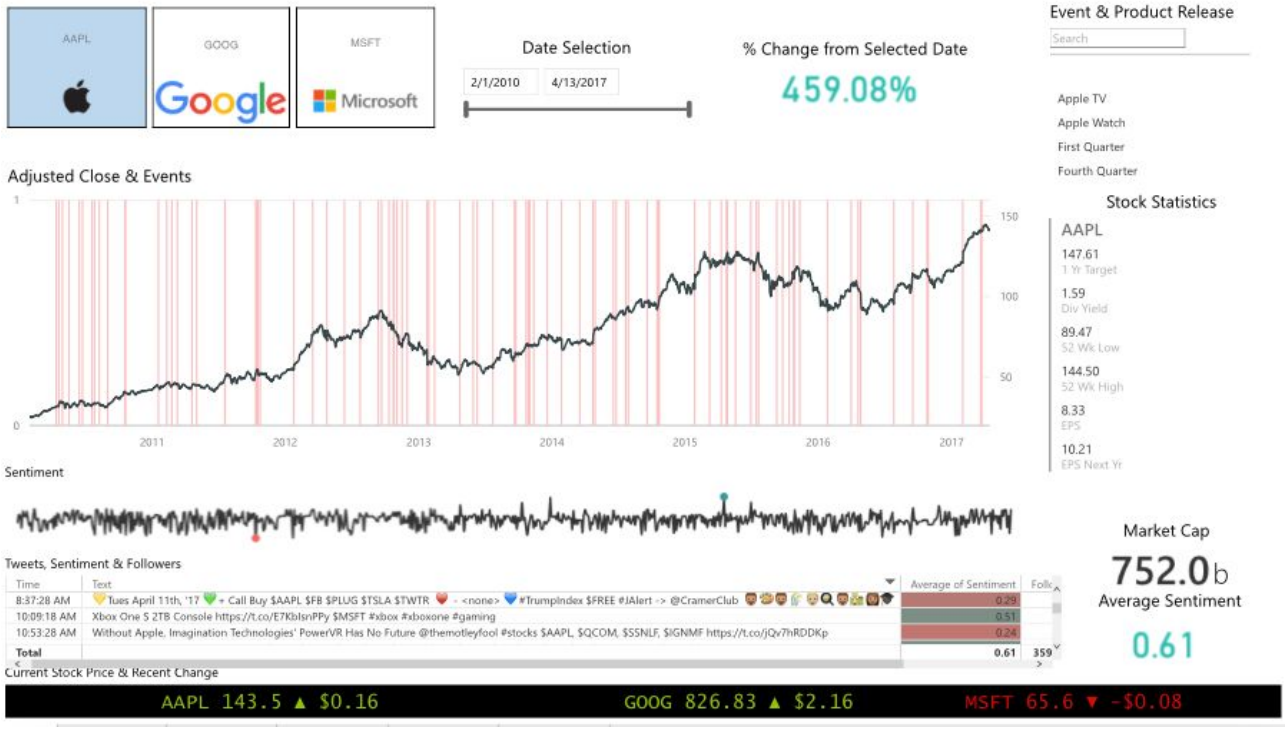
https://app.powerbi.com/visuals/. Among the ones we tested were Stock Chart, Pulse Chart, Hierarchy Slicer, Time Brush, Gap Analysis, and Cluster Map. One custom visualization that made it into our final dashboard was Chiclet Slicer, which allows the company's logos to be on the slicer instead of boring text. We also kept Attribute Slicer which allows users to select what main events or earnings releases that they want to show on the graph. Scroller is another custom visualization that is seen at the bottom of our dashboard and shows users the last trading price along with the change from the previous trade price. Sparkline was used to show the sentiment over time and is below the adjusted closing price graph with the high and low points in sentiment over time. Finally, the custom visualization Card with States is used for the percent change and sentiment cards because it allows us to show the text as green when it is positive or above .5 and red when it is below .5 or negative returns.

Other standard visualizations used in our dashboard is a combination line and bar graph which has the adjusted close prices as the line series and major events as the bar graph series. This was used because you can't have two time series on one chart in Power BI. Additionally, we have standard stock attributes in a card to the right of the stock chart to show our users standard info. Below the stock chart is a table of tweets along with the number of followers and their sentiment score. Above the stock chart graph are slicers and filters including a date slicer which allows you to select what dates of data you want to see. Finally, on the bottom right we have two cards showing the sentiment score of the day, and the market cap that is auto updated.

Exploratory Data Analysis

From the beginning, we had the idea of having our page show one company at a time, however, this developed into showing the trends of multiple companies based on which industry slicer the user selected. We wanted to show the trends of each company on the same chart, but the difference in stock prices loses integrity and isn't helpful to user at a glance. Thus we headed towards a dashboard that had individual stats (open, close, high, etc) for itself, then based on what the selected company was apart of, graphs comparing volume traded as well as stacked

trend lines displaying any positive/negative change. There is also a date feature that can be changed to show any time period starting from 2010.



Power BI File Name (WEB) - Published version, however can only update in desktop mode
https://app.powerbi.com/view?r=eyJrIjoiOTEyNTM5YzctYTJlYi00NGNkLWFiNGYtNDg2MjZjYjYyNzQ4IiwidCI6ImYzOWU1MjI0LTE1OTYtNGFjZS04NTNiLWYxZmY5YzI0NDgxZCIsImMiOjN9

Project Improvement for Final Submission

After getting feedback on our Project 1, we decided to improve our analytics and comparisons within the dashboard. First, we added the NASDAQ change in the market for the selected time period. This shows how the selected company performed when compared to the total tech market performance. We also improved our data scraping from Yahoo finance by having the data update real time when power BI is refreshed as well as consolidated data into power Bi charts instead of connecting imported excel files. We also included all the products and quarterly earnings for each of the three companies we were comparing. Lastly, we cleaned up the dashboard to improve the appearance and comparisons between graphs/charts.

Evaluation

Main takeaways from our project:

- Power BI is a powerful tool but lacks some key functionality including having two time series on one chart.
- Google and the Power BI forums are the best resources ever invented and helped us beyond belief with various problems.
- While it may seem like you can easily pull data from anywhere, it is actually pretty hard to source Twitter data among other data sources.
- We gained a new insight into how AAPL, MSFT, and GOOG stock fluctuates based on product releases and quarterly releases
- We believe that social influence from Twitter could influence how a stock goes up and down
- Excited to look at hout our company's compared to the NASDAQ

Previous Project Work & Schedule

Sunday, March 5th - Created questions and purpose for the project goal

Wednesday, March 15th - Developed sketches and plans for final dashboard including simple charts in a Tableau dashboard. These included slicers on specific companies and dates, as well as graphs showing negative or positive change per day, quarterly increases or decreases, volume by company, and highs of each showing trends. Our next goal is to implement an overlay showing events or announcements that are related to each industry. These images are below:

Saturday, April 8th - We decided to switch to Power BI for ease of use dashboard creation. Our main objective for project submission is to create a dashboard that is able to be filter by time range and company as well as show dates that could affect the stock. We also found a new way to pull data from Yahoo finance and then have it update daily for real time use, which we will include for project 2 submission. We also decided to cut our scope down and only focus on a few companies within certain industries.

Sunday April 9th - Focused on finalizing a page for the technology companies. Added a ticker and polished the process book.

Monday April 10 - Worked on project 1 further by implementing possible data of product releases and quarterly reports from each company. Enhanced process book and cleaned up project

Tuesday April 11 & Thursday April 13 - Met with Fadel to see where we were and understand what the deliverables were for Project 1, video explanation, and process book outline

Thursday April 13 - Completed product release overlay feature, added sentiment line graph, and cleaned up report

Monday April 17 - Finished up process book, finished draft for project 1 Power BI, updated manually and published to web. Recorded features and how to use them for a potential investor.

Sunday May 7th - Redid the Power BI notebook to include product releases from all three of our companies. Corrected and uploaded an HTML and PDF of process book to Github. Adding NASDAQ baseline and adjusting graphs for visual comparison. Cleaned up queries so the BI doesn't have an extended run time, and including real time updating for accurate data.