n

STOCK MARKET DATA VISUALIZATION AND FORCASTING USING TABLEAU

Sushant Burde

DATASET

High-quality financial data of US based stocks and ETFs trading on the NASDAQ MKT. A stock market, equity market or share market is the aggregation of buyers and sellers (a loose network of economic transactions, not a physical facility or discrete entity) of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange, as well as stock that is only traded privately.

STOCK MARKET DATA VISUALIZATION AND FORCASTING USING TABLEAU

ABSTRACT

In finance, market data is price and trade-related data for a financial instrument reported by a trading venue such as a stock exchange. Market data allows traders and investors to know the latest price and see historical trends for instruments such as equities, fixed-income products, derivatives, and currencies.

Tableau visualizes and analyzes millions of rows of stock data effortlessly. It was built without any programming by a regular business user. Tableau empowers people to ask and answer their own questions of the data at the speed of thought. Digging deeper into the data is as easy as dragging and dropping, pointing and clicking.

Open-High-Low-Close attributes used to illustrate movements in the price of a financial instrument over time. High and low price indicates the highest and lowest stock price on the date whereas open and close indicates the opening and closing price on the same day.

A candlestick is a type of price chart used that displays the high, low, open and closing prices of a security for a specific period. It originated from Japanese rice merchants and traders to track market prices and daily momentum hundreds of years before becoming popularized in the United States. The wide part of the candlestick is called the "real body" and tells investors whether the closing price was higher or lower than the opening price (black/red if the stock closed lower, white/green if the stock closed higher).

Keywords: Candlestick Charts, Stock Open/High & Close/Open Data, Tableau, Butterfly Chart, Lollipop Chart, Line and Bar Charts.

SAMPLE DATA

ticker	date	open	high	low	close	vol
AAME	20030101	1.63	1.63	1.63	1.63	0
AAON	20030101	3.6423	3.6423	3.6423	3.6423	0
AAPL	20030101	1.0236	1.0236	1.0236	1.0236	0
ABAX	20030101	3.96	3.96	3.96	3.96	0
ABCB	20030101	10.7917	10.7917	10.7917	10.7917	0
ABCO	20030101	14.95	14.95	14.95	14.95	0
ABMD	20030101	3.64	3.64	3.64	3.64	0
ACET	20030101	4.7319	4.7319	4.7319	4.7319	0
ACGL	20030101	10.39	10.39	10.39	10.39	0
ACLS	20030101	22.436	22.436	22.436	22.436	0
ACMR	20030101	12.71	12.71	12.71	12.71	0
ACTG	20030101	2.41	2.41	2.41	2.41	0
ACXM	20030101	15.38	15.38	15.38	15.38	0
ADBE	20030101	12.4005	12.4005	12.4005	12.4005	0
ADI	20030101	23.87	23.87	23.87	23.87	0
ADP	20030101	34.46	34.46	34.46	34.46	0
ADRA	20030101	15.651	15.651	15.651	15.651	0
ADRD	20030101	14.785	14.785	14.785	14.785	0
ADRE	20030101	12.75	12.75	12.75	12.75	0
ADRU	20030101	14.819	14.819	14.819	14.819	0
ADSK	20030101	7.15	7.15	7.15	7.15	0

Stock data always comprises of company tickers and date columns. High and low prices are the companies highest and lowest price till the trading runs for the day. Opening and closing price indicates the opening trading price and closing price shows the end price for that day. Volume column shows the stock quantity for each and every company within that day.

There are many different types of stock charts: line, bar, OHLC (open-high-low-close), candlestick, mountain, point-and-figure, and others, which are viewable in different time frames: most commonly, daily, weekly, monthly, and intraday charts. Each style and time frame has its advantages and disadvantages, but they all reveal valuable price and volume information that you can use to make profitable investing decisions.

STOCK MARKET DATA VISUALIZATION AND FORCASTING USING TABLEAU

DATA CLEANING

57749	SASR	20070309	33.69	34.05	3
57750	SBAC	20070309	27.2	27.43	2
57751	SBBX	20070309	15.5	15.5	
57752	SBBX	20070309	117.5	118.3	11
57753	SBBX	20070309	14.58	14.68	1
57754	SBNY	20070309	30.33	30.33	2
57755	SBSI	20070309	22.5238	22.581	22.(
57756	SBUX	20070309	15.46	15.495	15

Ticker columns has many duplicate company ticker for same date.

57750	SBAC	20070309	27.2	27.43	2
57751	SBBX	20070309	15.5	15.5	
57752	SBCF	20070309	117.5	118.3	11
57753	SBGI	20070309	14.58	14.68	1
57754	SBNY	20070309	30.33	30.33	2
57755	SBSI	20070309	22.5238	22.581	22.(
57756	SBUX	20070309	15.46	15.495	15
57757	SCHL	20070309	34.68	34.74	3

The companies ticker compared with other date to get the exact next ticker for same sequence.

591844	ONVI	20081014	4.25	4.27	4.25	4.27
591845	OPHC	##########	222.8	223.6	222.8	223.6
591846	OPNT	20081014	11.51	11.79	9.81	10.19
591847	OPOF	20081014	18.98	20.82	18.97	20.5
591848	OPTT	20081014	72.6	76	70	70
591849	ORBC	20081014	4	4.03	3.28	3.38
591850	ORBK	20081014	4.63	4.67	4.28	4.37
591851	OREX	20081014	78.6	80.5	71	74.3
591852	ORIT	20081014	17	17	15.56	16.27
591853	ORLY	##########	24.42	24.97	22.6	23.14
591854	OSBC	20081014	20.9	20.9	18.76	19.81
591855	OSBCP	20081014	8.6	9.4	8.05	9.31

Some Date Columns has date format issue

591843	ONEQ	20081014	
591844	ONVI	20081014	
591845	OPHC	##########	
591846	OPNT	20081014	
591847	OPOF	20081014	
591848	OPTT	20081014	
591849	ORBC	20081014	
591850	ORBK	20081014	
591851	OREX	20081014	
591852	ORIT	20081014	
591853	ORLY	##########	

Category: General, Number, Currency, Accounting, Date, Time, Percentage, Fraction, Scientific, Text, Special, Custom

Sample: #########

Type: *3/14/2001, *Wednesday, M, 3/14, 3/14/01, 03/14/01, 14-Mar, 14-Mar-01

Locale (location): English (U.S.)

Use the format cell option to convert the date format to YYYYMMDD.

923853	WAYN	20090918	5.75	5.75	5.75	5.75	0
923854	WDC	20090918	36.75	36.91	36.12	36.22	3884700
923855	WDFC	20090918	28.6	29.03	28.06	28.77	102100
923856	WERN	20090918	???	19.12	18.63	##	1073100
923857	WEYS	20090918	22.41	22.51	22.25	22.3	28900
923858	WILC	20090918	4.17	??	4.1	4.1	1500
923859	WINA	20090918	21.7	21.7	20.4	20.4	5500
923860	WIRE	20090918	24.25	24.28	24.03	24.24	193400

Bad character found in some columns for high low prices

923853	WAYN	20090918	5.75	5.75	5.75	5.75	0
923854	WDC	20090918	36.75	36.91	36.12	36.22	3884700
923855	WDFC	20090918	28.6	29.03	28.06	28.77	102100
923856	WERN	20090918	18.9	19.12	18.63	19.07	1073100
923857	WEYS	20090918	22.41	22.51	22.25	22.3	28900
923858	WILC	20090918	4.17	4	4.1	4.1	1500
923859	WINA	20090918	21.7	21.7	20.4	20.4	5500
923860	WIRE	20090918	24.25	24.28	24.03	24.24	193400

Filled missing columns with comparitive data as per the previous values. Bad character can be filled with relavtive values to avoid issue in visualization

VISUALIZATIONS

1. **Candlestick chart for Google in the period from 2007 to 2009.**

Fig. 1.a. Candlestick Chart for Google in the period Jun 22, 2008 to Aug 21, 2008

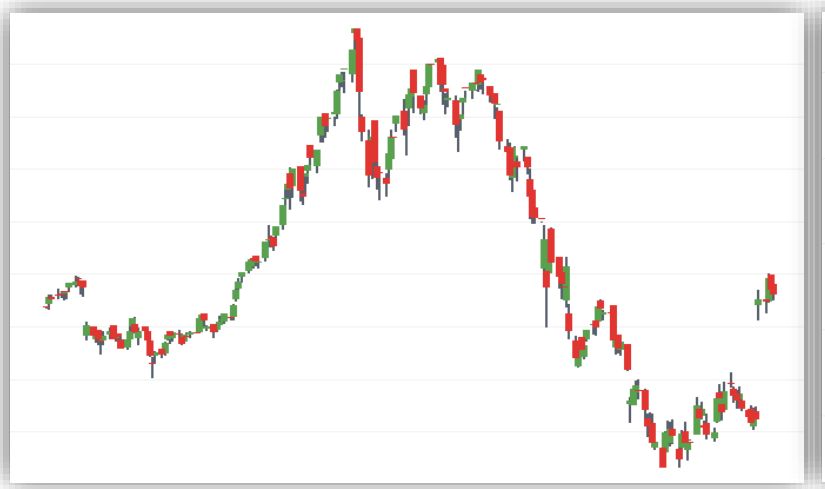


Fig. 1.b. Extended view of Candlestick chart

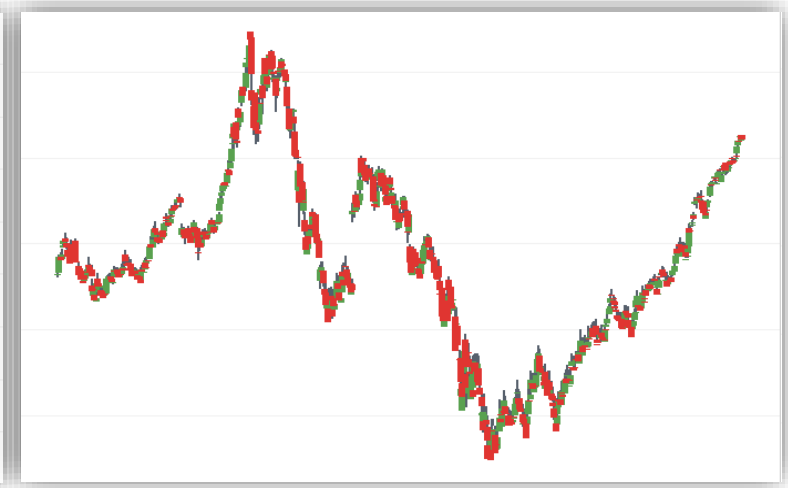


Fig. 1.c. Fully Enlarged view of Candlestick Chart

Candlestick charts can be created by **open, high, low and close** values for each time period. The hollow or filled portion of the candlestick is called “the body” (also referred to as “the real body”). The long thin lines above and below the body represent the high/low range and are called “shadows” (also referred to as “wicks” and “tails”). The high is marked by the top of the upper shadow and the low by the bottom of the lower shadow. If the stock closes higher than its opening price, a hollow candlestick is drawn with the bottom of the body represents the opening price and the top of the body representing the closing price. If the stock closes lower than its opening price, a filled candlestick is drawn with the top of the body representing the opening price and the bottom of the body representing the closing price.

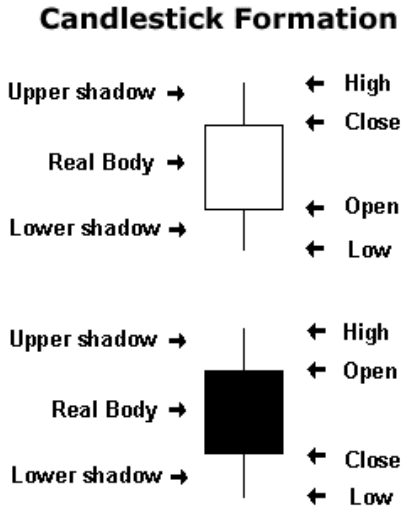


Fig.1.a. demonstrates the candlestick visualization for Google in an unexpended view for period Jun 22, 2008 to Aug 21, 2008. Green bar indicates that the opening price for stock is less than closing price. Similarly, red bar indicates the opening price for stock is greater than closing price. The close-open price difference shows the negative and position values which similarly defines the profit and loss for that particular day.

Fig 1.b. visualizes the extended view of the same stock to get the overall behavior. More extended view showcases the uptrend and downtrend of the stock.

Fig 1.c. Fully enlarged view for the period of 2007 to 2009. Google stock dimensionality has varied significantly during the period. In December 2007, the stock prices for Google increased highly towards the greatest peak. Later 1 year, around December 2008, stock prices dropped to its lowest peak.

2. Stock Comparison for Amazon, Google and Apple in the period of 2006 to 2009.

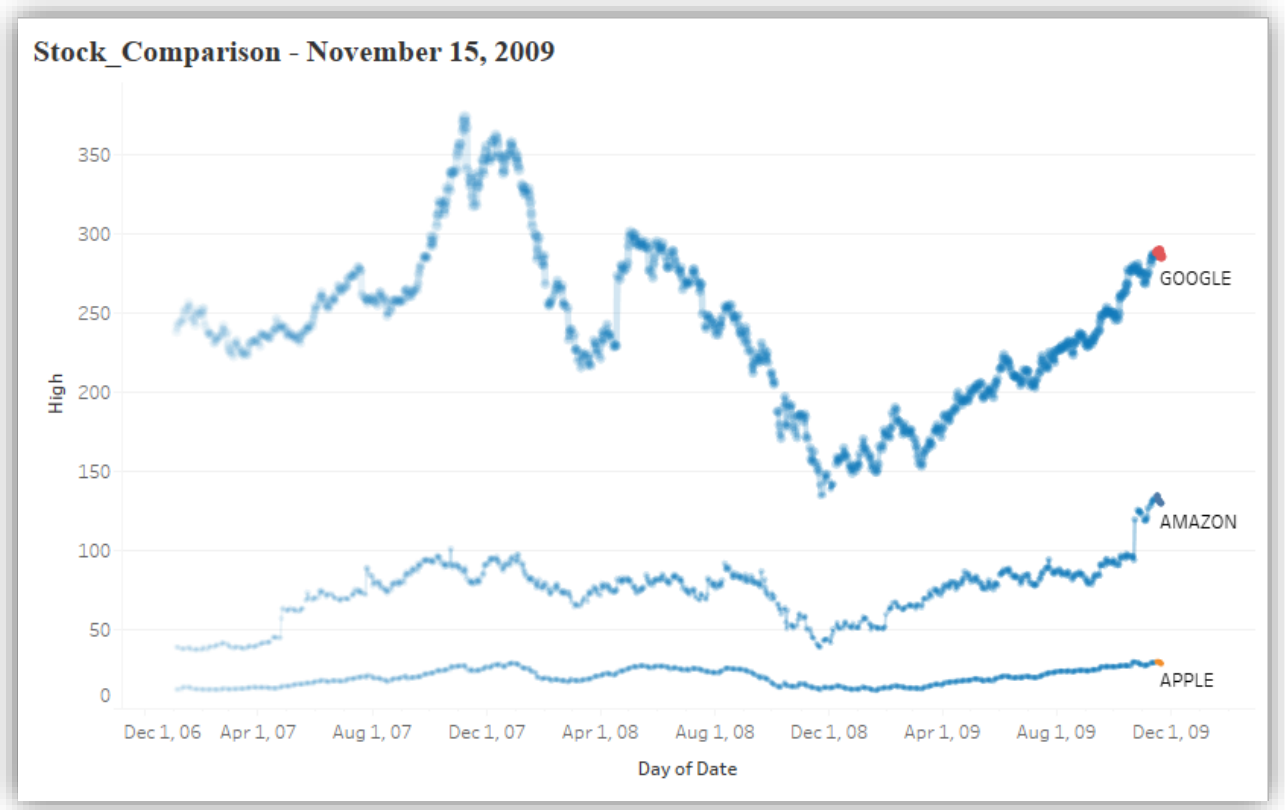


Fig.2. visualizes the stock comparison for three big companies in the period from 2007 to 2009. The internet giant Google started with less variation till August, 2007. The Rapid increase was shown after the consistent performance and touches to its highest peak producing huge profit. Later around April 2008, there is decline in the stock prices which produced a huge impact on the share market value. No growth recorded later and faced lot of inconsistency during the period. The dropped in share prices continues and recorded every time low price. The phase from December 2008 to December 2009 grows as the share market prices seek towards end of the year.

Amazon stocks performed significantly well and managed to stabilize the stock within the whole period of span. In figure shows, the deflection in the stock trends around December 2008 but later, no decline recorded. However there is a great uptrend recorded somewhere around November 2009. During the period, Amazon sold large number of products and produced magnificent sales profit. Comparing the stock prices from December 2006 to December 2009, high price value was reached to 140 Billion. Apple managed to provide stable growth, generate consistent revenue during the period.

3. Lollipop chart for Top 10 Companies as per the Adj. High Price.

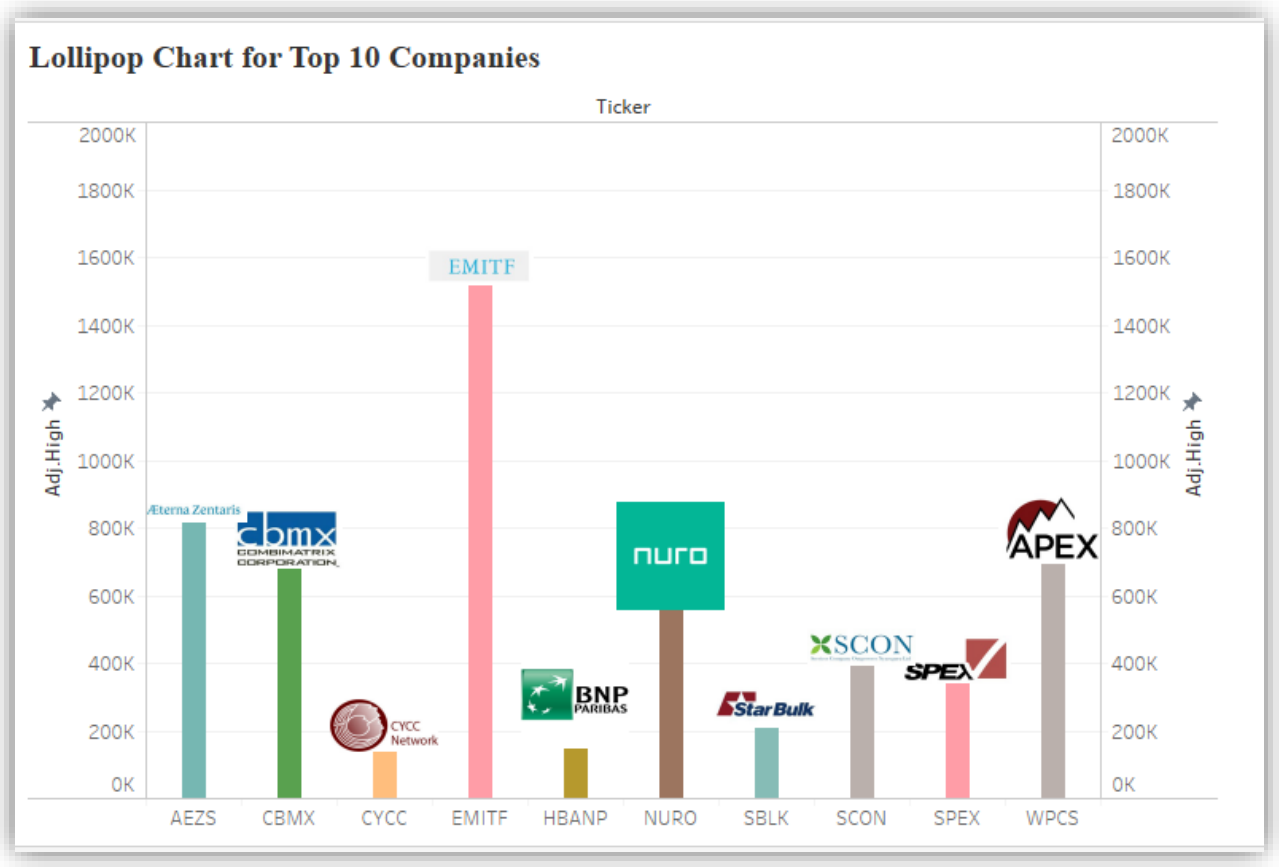


Fig. demonstrates the top10 companies based on their highest adj. High price. Lollipop chart is the combination of Dual Axis chart and circle chart. It is the best visualization to showcase the measurement of any dimension. The icons on the top create a good impression showing the proper stock companies completing in the market with other competitors. As shown in fig, EMITF company has made a significant growth and recorded the highest high stock price. CBMX and NURO managed an average stock price compared with other stock market competitors. CYCC Network was recorded low as compared to other stock companies during the entire phase.

\

4. **Butterfly Chart for Top 20 Companies indicating High-Low and Close-Open Difference.**

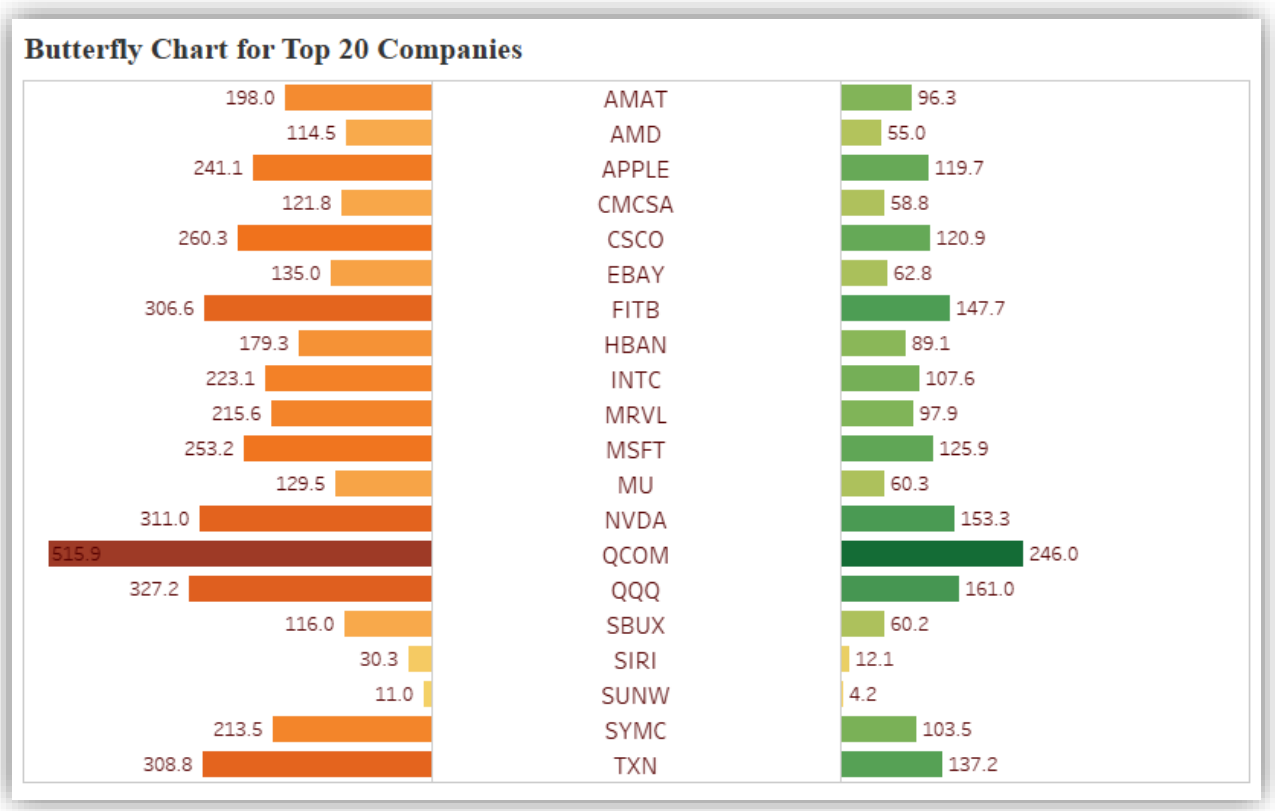
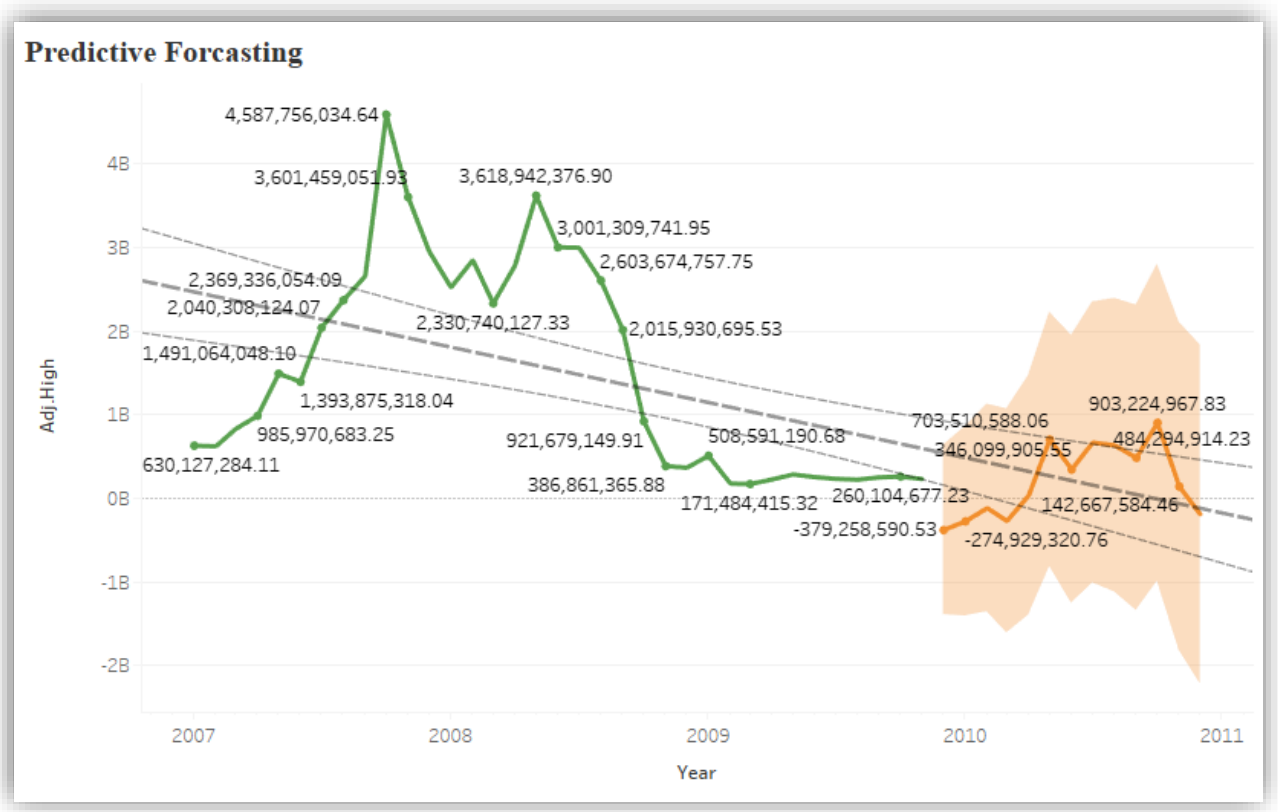


Fig. demonstrates butterfly chart which are in essence two bar charts on either side of the same y-axis. This chart helps you to easily portray Data with Dual Scales but having the same axis titles. The one side shows the high-low difference price for particular company. Similarly you can see the Open-Close difference simultaneously and compare the price range with each other. The visualization is created based on top 20 companies in US stock market. Qualcomm has made a significant impact in stock market with highest difference around 515 Million Dollars. Colors Marks has been used to show case the growth rate. Fade colors are used to show the less difference in high-low price. Sunworks stocks impacted with declined in their revenue and recorded all time low prices compared with other companies in the market.

5. **Predictive Forecasting for year 2010-2011 as per the stock available for 2007 to 2009.**

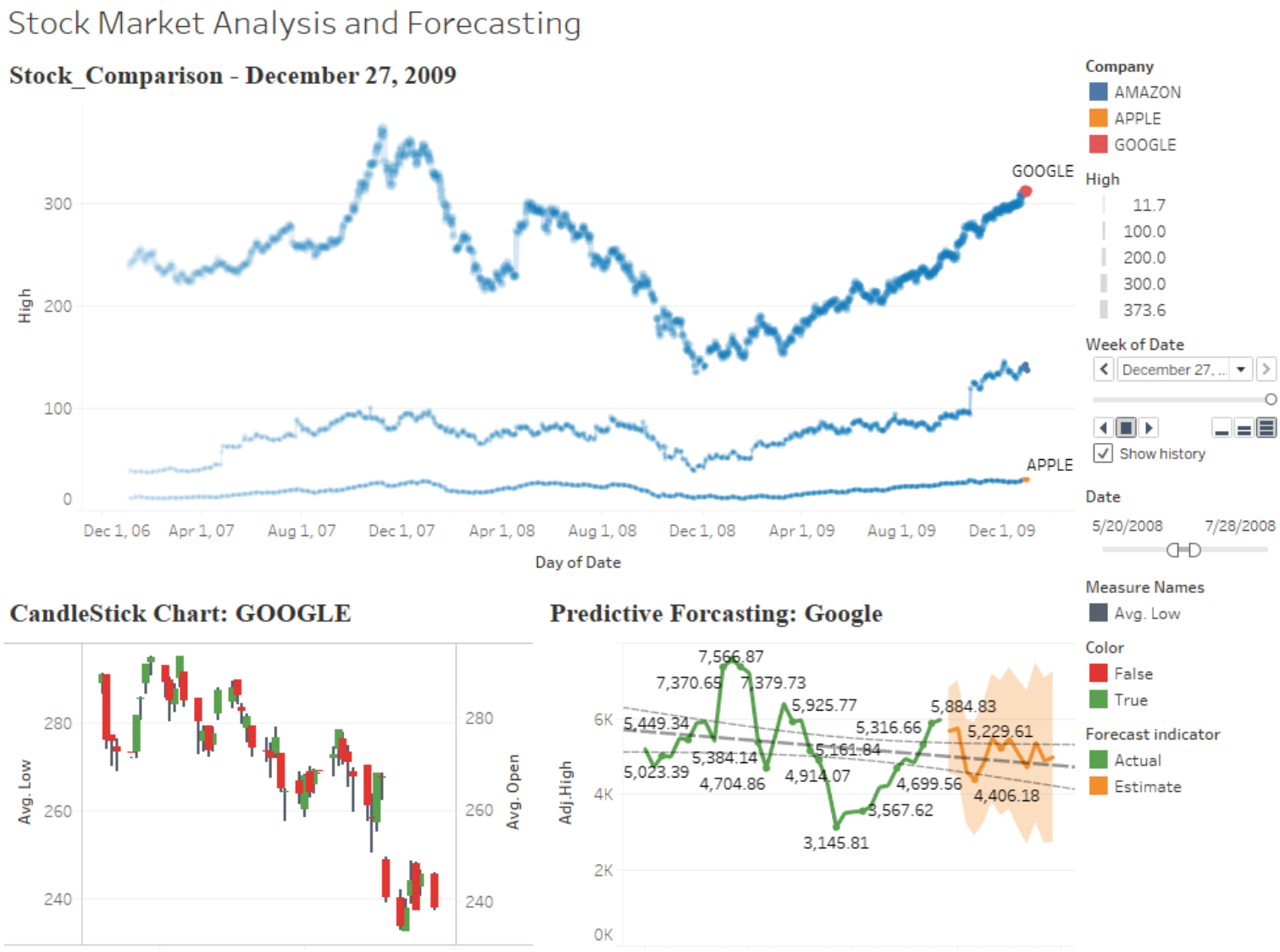


The above visualization shows the representation of the aggregated sales price for all the companies in the year 2007 to 2009. Forecast algorithms try to find a regular pattern in measures that can be continued into the future. It is also known as exponential smoothing. Tableau helps to create prediction based on the data available and forecast visuals as per the tolerance percent band. The green line graph shows the actual data and orange line indicates predicted values. The faded region dignifies the level of the prediction goes but not cross the edge of the band. Prediction level categories are 90, 95 and 99 % tolerance.

The middle line shows the best fit line which signifies the best equation to identify whether the X values is best for relevant Y value. R squared parameter is used to define how accurate the line fits in the graph. The accuracy level ranges from 0 to 1. More closely the R square towards the 1, more accurate you results will be. On average, 70 – 80% R square values considered best for any best fit line.

STOCK MARKET DATA VISUALIZATION AND FORCASTING USING TABLEAU

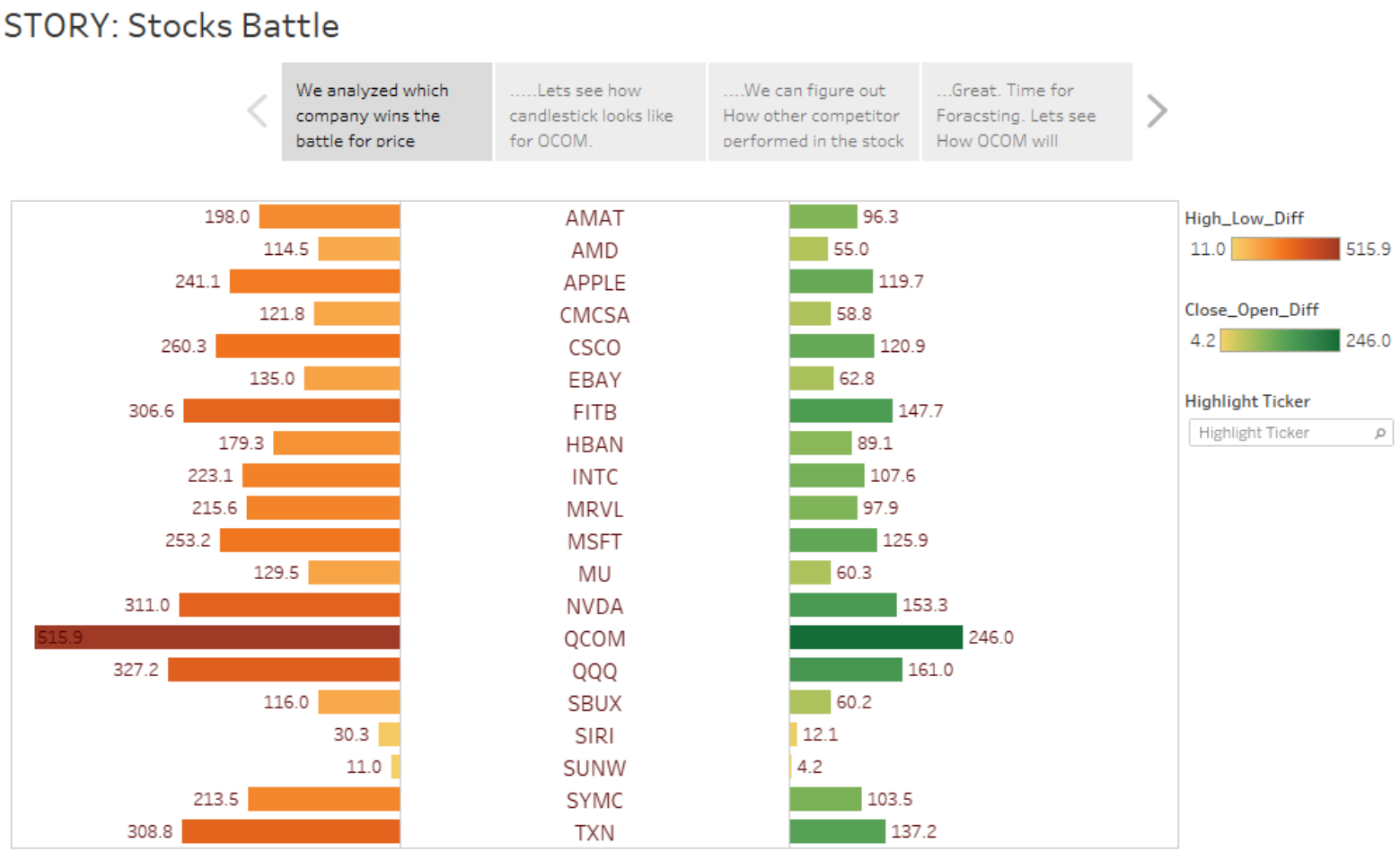
DASHBOARD: STOCK MARKET ANALYSIS FOR GOOGLE



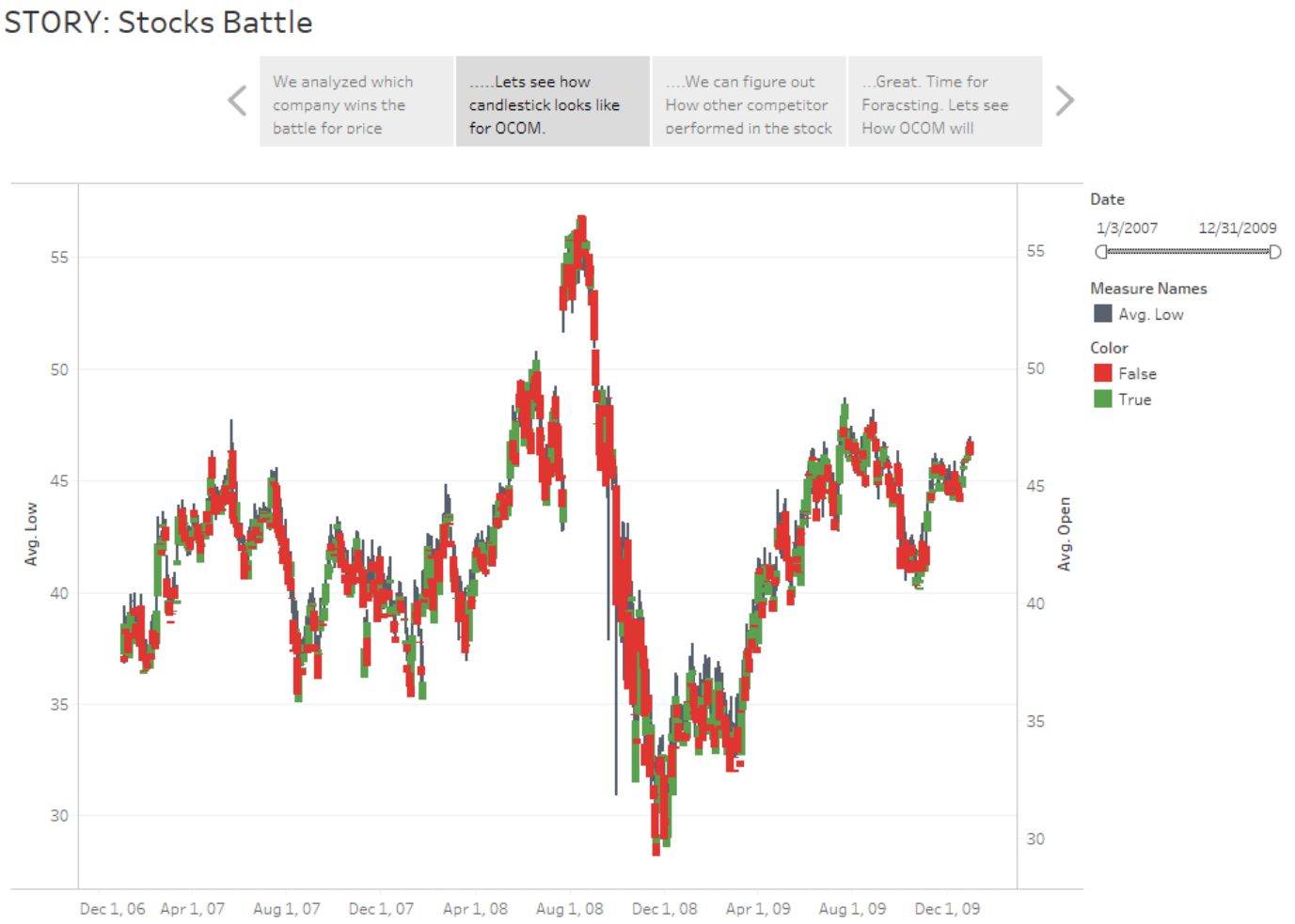
The dashboard demonstrates the stock market variation for internet giant Google. The market growth for Google is becoming increasingly reliant on its own sites to generate cash: two thirds of revenues were generated on Google-owned properties, with the remaining third generated by the company's extensive network of advertising partners. You can simultaneously visualize the stock trends for other companies and see the movements of a security, derivative, or currency from the candlestick diagram for every date in the period of 2007 to 2009.

We can also judge the performance will be for next year. It will help customers or stock traders to invest money and gives ideas about the perfect trade day. Tolerance band will help us to understand fluctuation of stock spikes. Uptrends and downtrends can be easily understandable using these charts.

STORY: STOCK PRICE BATTLE

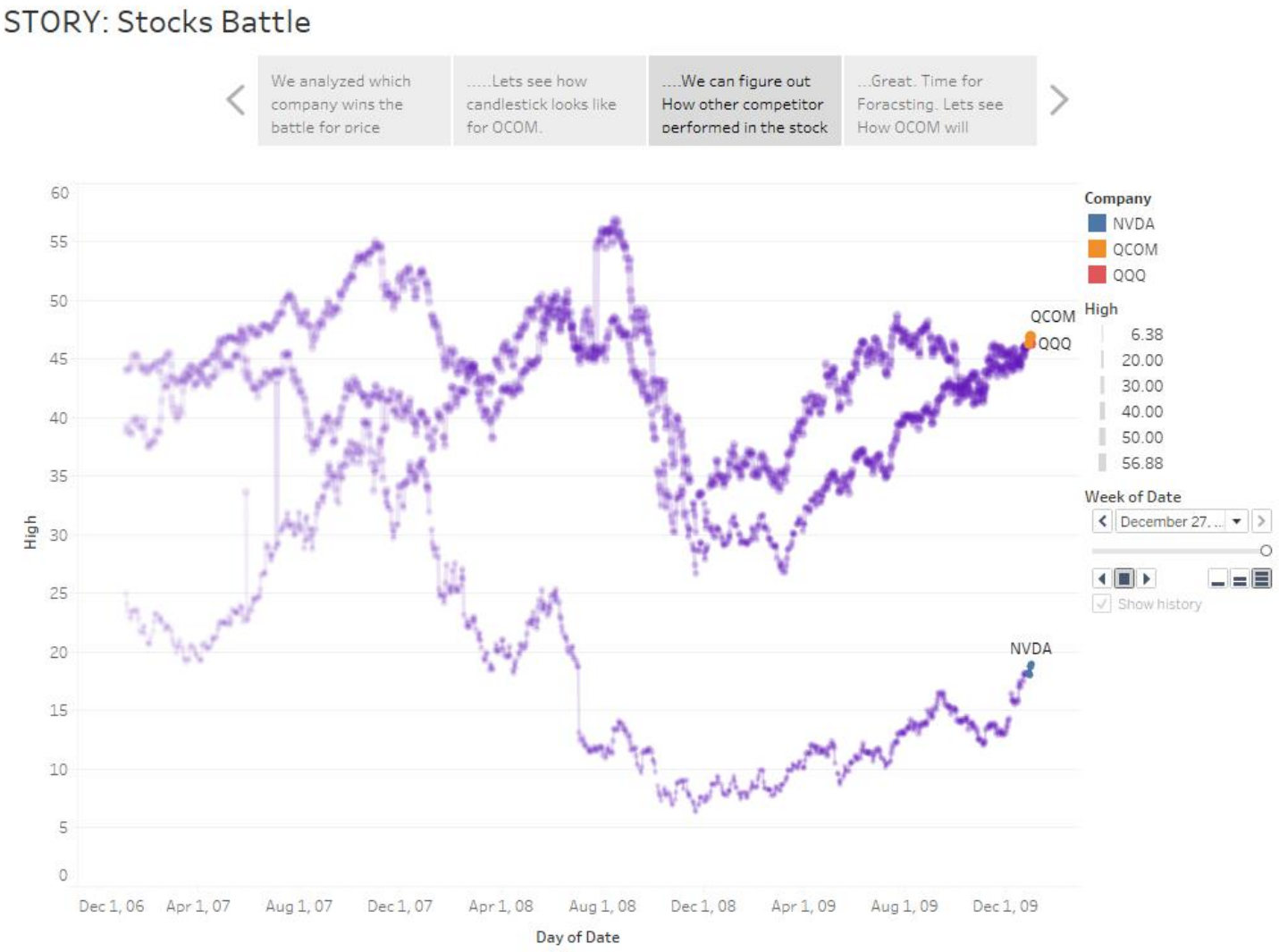


The QUALCOMM wins the battle for price difference in high-low and close-open price variation. Invesco QQQ and NVidia were also in the race and competing against the Qualcomm. In the next tab, we will look at the one day's worth of price data about a stock. A candlestick recognizable pattern helps investors to make buying and selling decisions. Doji and Harami pattern will be useful to see the each day price comparison.



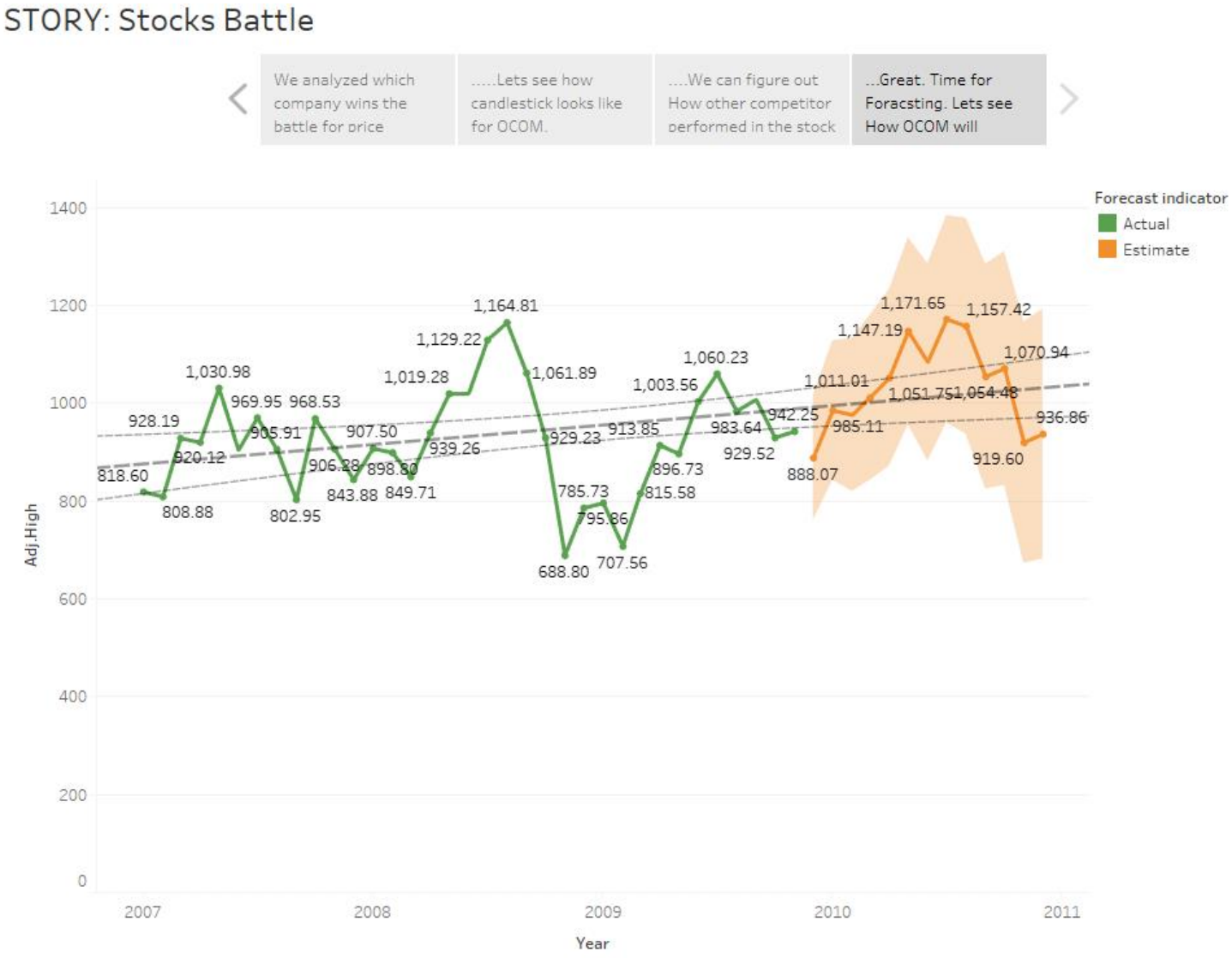
Dozens of bullish and bearish live candlestick chart patterns for the Qualcomm Inc. BDR stock. QUALCOMM Incorporated candlestick chart analysis, stock chart patterns with Fibonacci retracement lines. The Qualcomm stock patterns are available in a variety of time frames for both long and short term investments. Qualcomm stock has lots of Hammer and Hanging Man chart which looks exactly alike, but have different implications based on the preceding price action.

Dragonfly doji form as also shown in Qualcomm when the open, high and close are equal and the low creates a long lower shadow. The resulting candlestick looks like a “T” due to the lack of an upper shadow. Dragonfly doji indicate that sellers dominated trading and drove prices lower during the session. By the end of the session, buyers resurfaced and pushed prices back to the opening level and the session high.



As shown in fig, Qualcomm and QQQ battling each other throughout the period from 2007 to 2009. The overlapping took place during March 2008 and September 2008. From April 2007 to September 2008, Qualcomm revenues have increased remarkably well compared to QQQ. Simultaneously, the share prices for QQQ dropped and market price value decrease gradually.

NVidia started with low high price value and keep on decreasing through the years. The two spikes show the sudden increase in revenues for NVidia which increased the share price value. Company made lot of profit for those days. Now we looked how Qualcomm made a significant impact on stock price, we need to see where the market price goes for next year.



Finally, we can judge that Qualcomm will make the stock prices up towards the March or April 2010. The total prices for share price will goes around 1.17 Million Dollars which helps stock traders understand the amount to invest in this company. The minimum price predicted was around 0.8 Million Dollars. Also the best fit line will be helpful to keep an eye on the days and price value relation. Tolerance bands will be useful analytical concepts to judge the variation in the stock uptrends and downtrends.

REFERENCES

1. Understanding candlestick patterns and formation.
   https://stockcharts.com/school/doku.php?id=chart_school:chart_analysis:introduction_to_candlesticks
2. Butterfly chart explanation and concepts
   https://www.thedataschool.co.uk/will-griffiths/creating-butterfly-chart-tableau/
3. Importance of lollipop chart and its uses
   https://www.tableau.com/about/blog/2017/1/viz-whiz-when-use-lollipop-chart-and-how-build-one-64267
4. Line chart explanation
   https://www.tutorialspoint.com/tableau/tableau_line_chart.htm
5. Understanding stock charts and prices
   http://www.mit.edu/~mbarker/formula1/f1help/11-ch-10.htm
6. Real time stock trading
   https://www.daytrading.com/charts
7. Reference Stock Comparison charts
   https://www.nasdaq.com/aspx/mutual-fund-chart.aspx?symbol=NDAQ&selected=NDAQ