Elon’s tweet vs Tesla price change

December 14, 2021

Sean(Kwang Heum) Yeon

Contents

1. Introduction

1.1 Research question

How much does Elon Musk's number of tweets per day affect Tesla's stock price change?

1.2 Motivation

One day I met a friend at a social gathering. At the same age as me, a former lawyer introduced himself as a full-time father after retiring earlier this year.

- *Sean: How did you already retire at such an early age? Are you taking a break?*
- *Friend: Actually, not. I don't have to work anymore with the realized income from Tesla shares.*
- *Sean: It is still difficult for me to predict stock market. Do you have any secrets?*
- *Friend: Rather than relying on analysis reports, I watched and observed Elon's Tweeting. I figured that he tweets before they release favorable news to the public.*

After that, I observed Elon's pattern of tweets, and I could see the stock price fluctuating after he left several tweets a day over the past few months. Therefore, through this project, I would like to clear the question of 'How much does Elon Musk's number of tweets per day affect Tesla's stock price change?'

Elon Musk's daily tweets (Source: Twitter)

1.3 Why we should care about this question

For stock investors, especially day-traders, the rate of change in stock prices is an important factor for short-term planning. Even if the degree of change is small, it can be usefully reflected in the stock price prediction model if the variable has statistically convincing evidence.

2. Data Investigation Overview

We will apply simple linear regression to the analysis. Stock price change (%) = β0 + β1*tweet. We will gather two separate datasets, one for tweets and the other one for the stock price, and then combine those two. The period of the original dataset ranges between 2010 and 2021 and moving averages per lengths. With the datasets, we will statistically analyze them with a simple linear regression. Finally, we will interpret the output from different points of view.

As a disclaimer, our major question, "How much does Elon Musk's number of tweets per day affect Tesla's stock price change?" may have several different approaches to get the most trustful answer depending on each perspective.

2.1 Data sources

We referred to two sources - Tweet and stock price. Tweet dataset comes from 'Kaggle.com' (URL: https://www.kaggle.com/ayhmrba/elon-musk-tweets-2010-2021), and stock price dataset is based on 'Nasdaq.com' (URL: https://www.nasdaq.com/)

2.2 Data analysis tools

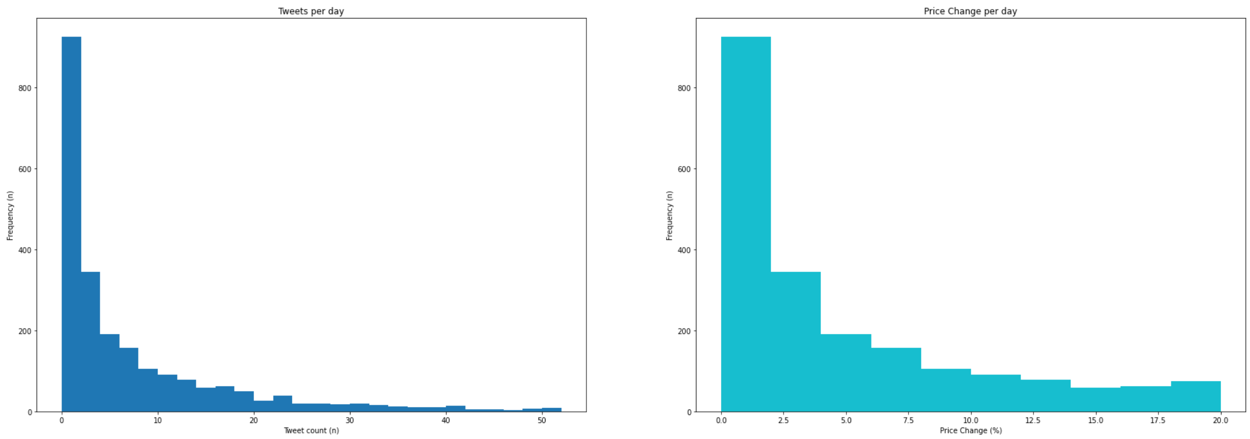
We used Microsoft Excel and Python to clean and organize those two datasets and as an analysis tool.

3. Data Summary

The original dataset includes stock price and Elon's tweets records between January 2010 and March 2021. Elon started his first tweet in 2010 and still using Twitter as the major social media. Because there was no tweeting until January 2012 after his first tweet in January 2010, we include datasets from January 2012 to March 2021. There were 24,056 tweets during the period. After grouping the tweets by date, we got 2,349 unique entries for one qualitative variable and two quantitative variables. The dependent variable in the analysis is 'price change (%)' and the independent variable is 'Tweet count.'

- Qualitative variable: date ('yyyy-mm-dd')
- Quantitative variable: price change (%), Tweet count per day (integer)

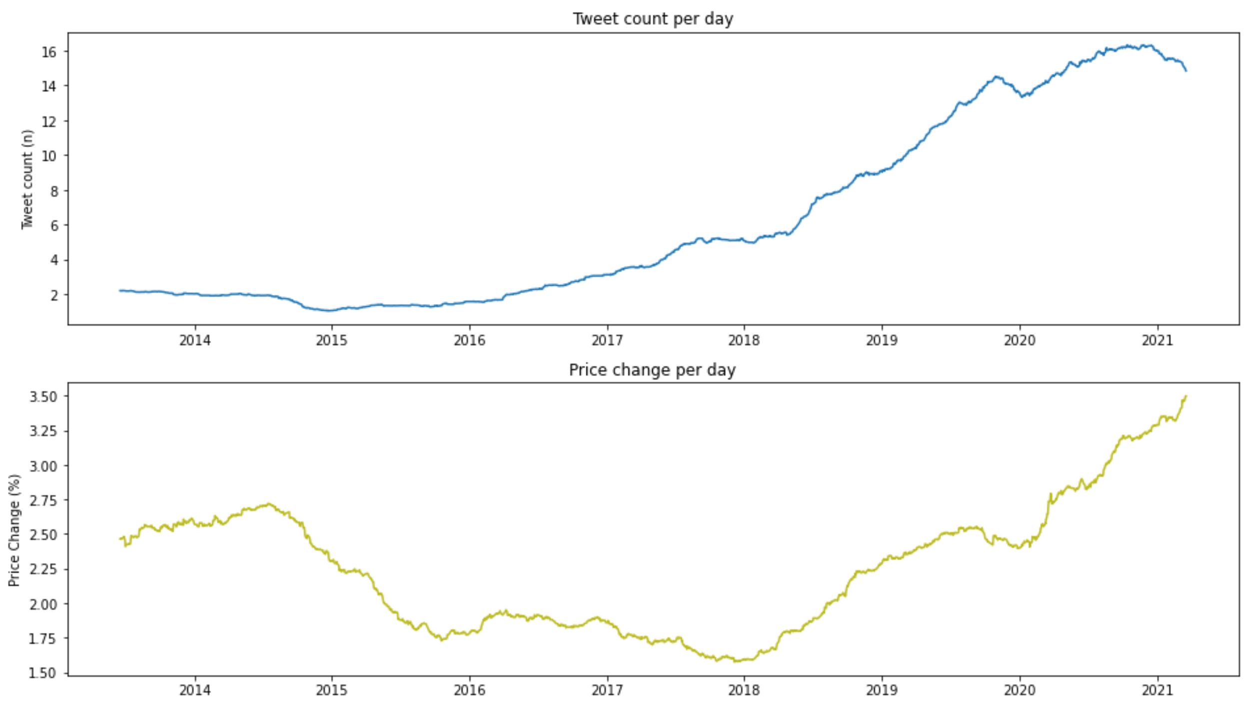
3.1 Histogram



Histogram for Tweet per day and Tesla price change (by Sean Yeon)

Both datasets show a right-skewed pattern with its mode around the beginning of datasets.

3.2 Line plot – Tweet vs Price Change



Tesla Stock price (Source: Nasdaq, Designed by Sean Yeon)

Comparing the two graphs above, it seems that the stock price has proportionally increased and decreased according to the number of tweets a day over the period.

3.3 Five-number Summary and Correlation coefficient

According to Investopedia, common moving average lengths are 10, 20, 50, 100 and 200.

10-day moving average

	price_change	tweet
count	2281.000000	2281.000000
mean	**2.367163**	**6.459097**
std	1.135270	6.578821
min	0.524296	0.000000
25%	1.586211	1.400000
50%	2.071912	3.600000
75%	2.837720	10.200000
max	9.511406	32.800000

Correlation coefficient: 0.347522

20-day moving average

	price_change	tweet
count	2271.000000	2271.000000
mean	2.356078	6.465918
std	0.974178	6.238348
min	0.736068	0.100000
25%	1.688797	1.500000
50%	2.117926	3.600000
75%	2.799763	10.400000
max	7.466539	27.300000

Correlation coefficient: 0.413212

50-day moving average

	price_change	tweet
count	2241.000000	2241.000000
mean	2.349229	6.484766
std	0.824967	5.966694
min	1.068783	0.480000
25%	1.742189	1.600000
50%	2.145273	3.920000
75%	2.793376	11.520000
max	5.938148	22.160000

Correlation coefficient: 0.509014

100-day moving average

	price_change	tweet
count	2191.000000	2191.000000
mean	**2.343527**	**6.508772**
std	0.725674	5.831767
min	1.370850	0.800000
25%	1.770261	1.580000
50%	2.198515	3.880000
75%	2.772234	12.640000
max	4.577250	19.980000

Correlation coefficient: 0.560411

200-day moving average

	price_change	tweet
count	2091.000000	2091.000000
mean	2.311758	6.422233
std	0.623407	5.613473
min	1.439701	1.040000
25%	1.785463	1.620000
50%	2.191361	4.020000
75%	2.657111	12.570000
max	4.140452	18.330000

Correlation coefficient: 0.599541

365-day moving average

	price_change	tweet
count	1926.000000	1926.000000
mean	2.249040	6.176967
std	0.466759	5.231497
min	1.574526	1.052055
25%	1.834991	1.939726
50%	2.237647	3.586301
75%	2.561961	10.360274
max	3.498932	16.328767

Correlation coefficient: 0.584428

Five-number summary and correlation coefficient (by Sean Yeon)

In the moving average of "200-day", we have the highest correlation coefficient, 0.599. "10-day" and "100day" moving averages have the highest mean value for "price_change" and "tweet," respectively.

(continued)

7.1 10-day moving average

In [37]:
```python
mvp = 10
i = 0

mavg10 = pd.DataFrame()

mv_pchange = np.array(df.price_change)
price_change_mv = []
for i in range(len(mv_pchange)-mvp+1):
    price_change_mv.append(np.average(mv_pchange[i:mvp+i]))
    i+=1
mavg10['price_change'] = pd.DataFrame(price_change_mv)

mv_tweet = np.array(df.tweet)
tweet_mv = []
for i in range(len(mv_tweet)-mvp+1):
    tweet_mv.append(np.average(mv_tweet[i:mvp+i]))
    i+=1
mavg10['tweet'] = pd.DataFrame(tweet_mv)
```

In [38]:
```python
print(mavg10.describe())
```

```
       price_change        tweet
count   2281.000000  2281.000000
mean       2.367163     6.459097
std        1.135270     6.578821
min        0.524296     0.000000
25%        1.586211     1.400000
50%        2.071912     3.600000
75%        2.837720    10.200000
max        9.511406    32.800000
```

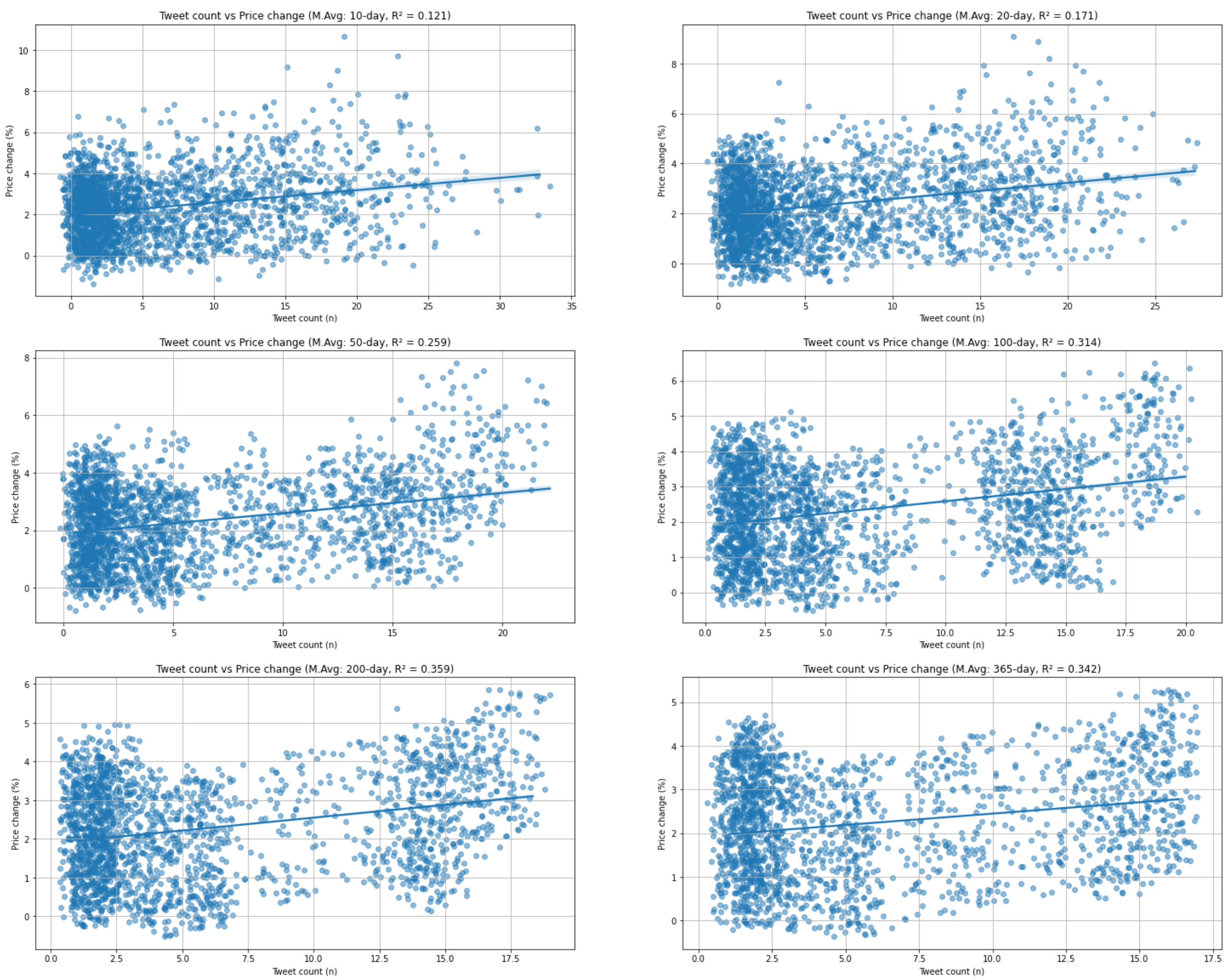
Code cells (Source: Jupyter Notebook by Sean Yeon)

Symbol	Last Price	Change	Chg %	Currency
NVDA	281.61	-20.37	-6.75%	USD
TSLA	966.41	-50.62	-4.98%	USD
PLTR	18.40	-0.54	-2.85%	USD
HOOD	19.70	-0.43	-2.14%	USD
AAPL	175.74	-3.71	-2.07%	USD
AMZN	3,391.35	-52.89	-1.54%	USD
GOOGL	2,916.53	-43.50	-1.47%	USD

Stock Price Change (Source: Yahoo Finance)

4. Inferences - Simple linear regression

For ‘Stock price change (%) = β0 + β1*tweet(n)’, we fitted each dataset into a linear regression model by OLS (ordinary least squares).



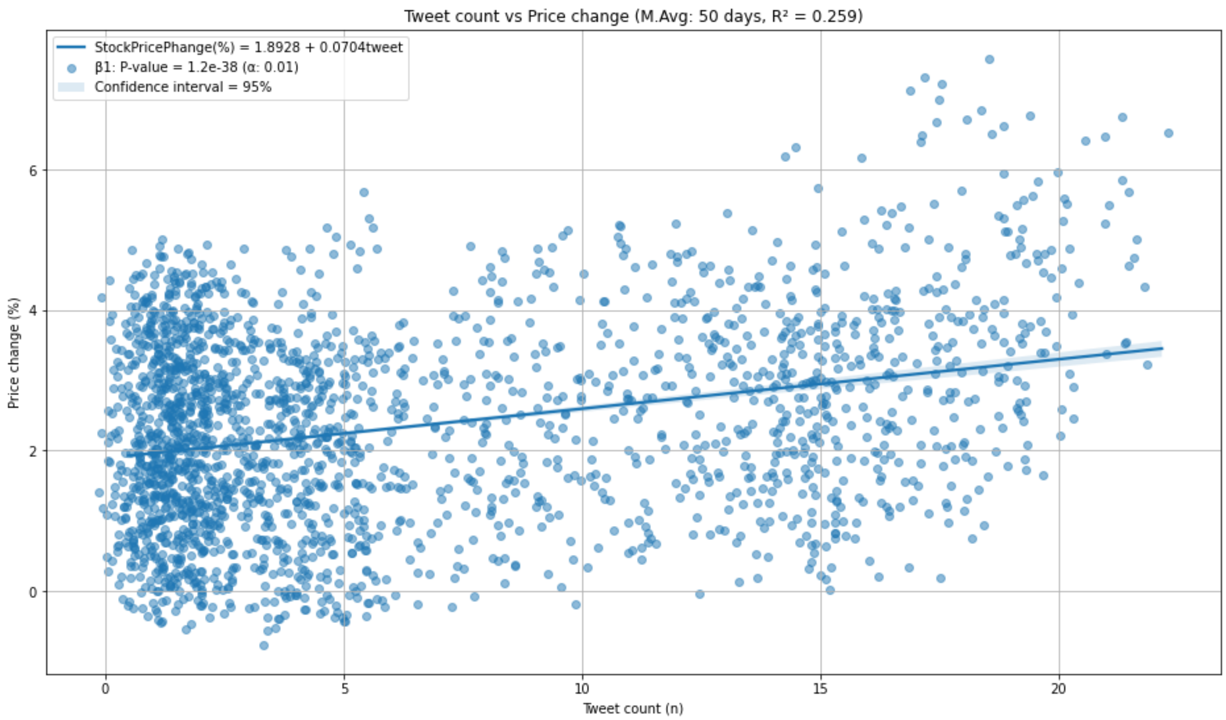
Simple linear regression model (by Sean Yeon)

- For ‘Stock price change(%) = β0 + β1*tweet(n)’,

  H0:β0 = 0 vs H1: β0 ≠ 0 and H0:β1 = 0 vs H1: β1 ≠ 0
- We assume our critical value(α) in this analysis at 0.01 or 1%.
- The p-value for the slope in each dataset has less than our critical value (0.01 or 1%), which suggests that the number of tweets in the datasets is statistically significant in relation to the stock price change.
- 200-day moving average has the highest R-squared, 0.359.

To find the most used moving average in the stock market, we referred to Investopedia, “Which moving average is the most important?” We will interpret the “50-day” moving average here according to the reference.

50-Day moving average



- The R-squared is 0.259, meaning that 25.9% of the variability in the response - stock price change is explained by the number of tweets through the linear regression model; in other words, the remaining 74.1% of the variability is due to characteristics that are not the number of tweets.
- ***Stock Price Change (%) = 1.8928 + 0.0704*tweet(n)***
- The predicted stock price change (%) is equal to 1.8928 plus 0.0704 times the number of tweets.
- If there is no tweeting, we would predict that the stock price has about 1.8928% change on average.
- Every one-unit increase in the number of tweets, we would predict that the stock price would change by 0.0704% on average.

5. Conclusion

We conducted simple linear regression and statistical analysis for the research question, 'How much does Elon Musk's number of tweets per day affect Tesla's stock price change?'. The R-squared in our model is 0.259, meaning that 25.9% of the variability in the response - stock price change is explained by the number of tweets through the linear regression model; in other words, the remaining 74.1% of the variability is due to characteristics that are not the number of tweets.

The predicted stock price change (%) equals 1.8928 plus 0.0704 times the number of tweets. If there is no tweeting, we would expect the stock price to have about 1.8928% change on average. For every one-unit increase in the number of tweets, we would predict that the stock price would change by 0.0704%. For example, if Elon tweets 20 times a day, it may trigger the stock price change about 1.8928 + 0.0704 x 20, or 3.30% on average in the 50-day moving average.

We also researched what a good R-squared value is. According to Investopedia, "What qualifies as a "good" R-Squared value will depend on the context. In some fields, such as the social sciences, even a relatively low R-Squared such as 0.5 could be considered relatively strong." Therefore, our R-squared analysis is worth considering because a person's tweeting trend is more closely related to the finance vs. social sciences than stock market indexes.

6. Reference

- *A good R-squared value ("R-Squared", Investopedia, Jason Fernado, Sep 12, 2021): https://www.investopedia.com/terms/r/r-squared.asp*
- *Small R-squared value ("Can a Regression Model with a Small R-squared Be Useful?", The Analysis Factor, Karen Grace-Martin, Apr 23, 2019): https://www.theanalysisfactor.com/small-r-squared/*
- *Elon Musk's tweets ("Elon Musk Tweets (2010 - 2021)", Kaggle, Ayhm Dalila, Mar 23, 2021): https://www.kaggle.com/ayhmrba/elon-musk-tweets-2010-2021)*
- *Tesla stock price ("Downloading historical stock prices in Python", toward data science, Rohan Joseph, Oct 13, 2020): https://towardsdatascience.com/downloading-historical-stock-prices-in-python-93f85f059c1f*
- *Useful moving average ("How to Use a Moving Average to Buy Stocks", Investopedia, Cory Mitchell, Apr 28, 2021): https://www.investopedia.com/articles/active-trading/052014/how-use-moving-average-buy-stocks.asp*
- *Most important moving average ("Why the 50-Day Simple Moving Average Is Popular Among Traders", Investopedia, J.B. MAVERICK): https://www.investopedia.com/ask/answers/012815/why-50-simple-moving-average-sma-so-common-traders-and-analysts.asp*

7. Source Code

https://github.com/cpasean/Projects/blob/main/Project_Business%20Analysis_TweetCount_vs_TeslaStockPriceChange.ipynb

December 14, 2021, by Sean Yeon