DA 460 / MDA 720 FINAL
Twitter Synthetic Control

Ari Bazigos, Mike Beccaris, Ryan Blume, Carter Ekberg, Tyler Welsh

Overview

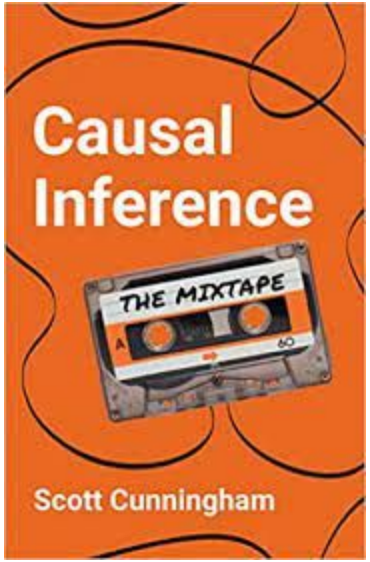
- Objective & Data Source
- Data Preprocessing
- Exploratory Analysis
- Synthetic Control for Sentiment Score
- Synthetic Control for Stock Price

Objectives

- Effectively scrape Twitter data
- Use VaderSentiment to calculate sentiment scores on tweets.
- Collect sentiments related to various social media platforms.
  - Included Facebook, LinkedIn, Instagram, Snapchat, and Reddit
- Observe the effect of Elon Musk buying Twitter on Sentiment scores using Synthetic Control
- Observe the effect of Elon Musk buying Twitter on Stock Price using Synthetic Control

References and Data Source

- Twitter (Data Source)
- Causal Inference: The Mixtape (Scott Cunningham)
- Causal Inference for The Brave and True (Matheus Facure Alves)



DATA PREPROCESSING

Data Preprocessing

- We first had to scrape Twitter for tweet contents regarding various keywords.
- Keywords included [Twitter, Facebook, Instagram, Snapchat, LinkedIn, and Reddit].
- Scraped 2,000 tweets per day for each platform from 4/19/22 to 4/30/22
- Elon Musk bought Twitter on 4/25/22, so we had a few days before and a few days after in order to test the effect of this treatment.
- We then combined all the CSVs created by the scraping function, and created a column to signify which social media platform the tweet was about.

	id	date	user	content	likes	retweets	quotes	replies	company
0	1516567316531748866	2022-04-19 23:59:59+00:00	https://twitter.com/Jazzboneplyr1	@Alex343 @eb454 @paulkrugman @Twitter we were speaking about harassment of people wearing masks	1	0	0	2	Twitter

Sentiment Scores

- After scraping and processing the data, we applied VaderSentiment to obtain sentiment scores for each of the 144,000(!) tweets in the dataframe.
- We then grouped by date and found the average positive, negative, and compound scores for each social media platform.
- This new dataframe allowed us to apply synthetic control.

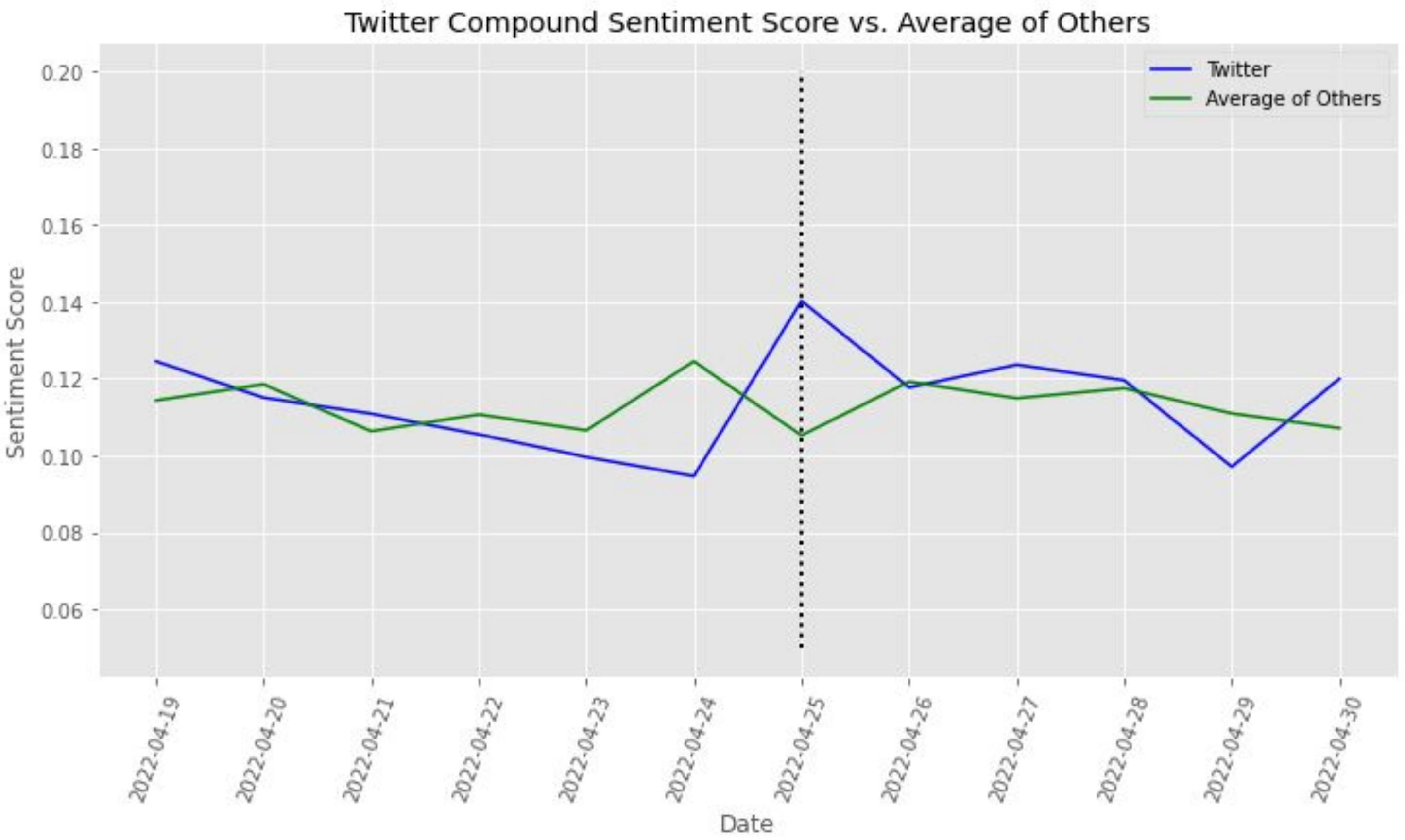
id	date	user	content	likes	retweets	quotes	replies	company	neg	neu	pos	compound
6531748866	2022-04-19 23:59:59+00:00	https://twitter.com/Jazzboneplyr1	@Alex343 @eb454 @paulkrugman @Twitter we were speaking about harassment of people wearing masks	1	0	0	2	Twitter	0.226	0.774	0.000	-0.5423
6292644864	2022-04-19 23:59:59+00:00	https://twitter.com/SandaraMary	@mhdksafa @Twitter Support your views!	0	0	0	0	Twitter	0.000	0.572	0.428	0.4574

EXPLORATORY ANALYSIS

Structuring Data for Plotting

company	Date	neg	pos	neu	compound
Facebook	2022-04-19	0.074993	0.101807	0.823210	0.109483
	2022-04-20	0.071859	0.099841	0.828300	0.104920
	2022-04-21	0.070761	0.105739	0.823495	0.117222
	2022-04-22	0.069179	0.107949	0.822871	0.123297
	2022-04-23	0.068549	0.107415	0.824033	0.127343
...	...	...	...	...	...
Twitter	2022-04-26	0.075836	0.116013	0.808158	0.117697
	2022-04-27	0.076698	0.118579	0.804717	0.123595
	2022-04-28	0.074909	0.118820	0.806273	0.119560
	2022-04-29	0.080138	0.115161	0.804706	0.097044
	2022-04-30	0.078179	0.118354	0.803474	0.119938

- Grouped all sentiment scores by ‘company’ and ‘Date’
- Used mean() to aggregate
- Split df into Twitter and all others
- Re-aggregated all others
- Plotted mean Twitter sentiment score per day against all others combined



Twitter Compound Sentiment Score vs. Average of Others
Twitter
Average of Others
Sentiment Score
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
2022-04-19
2022-04-20
2022-04-21
2022-04-22
2022-04-23
2022-04-24
2022-04-25
2022-04-26
2022-04-27
2022-04-28
2022-04-29
2022-04-30
Date

SYNTHETIC CONTROL FOR SENTIMENT SCORES

Initial Issues

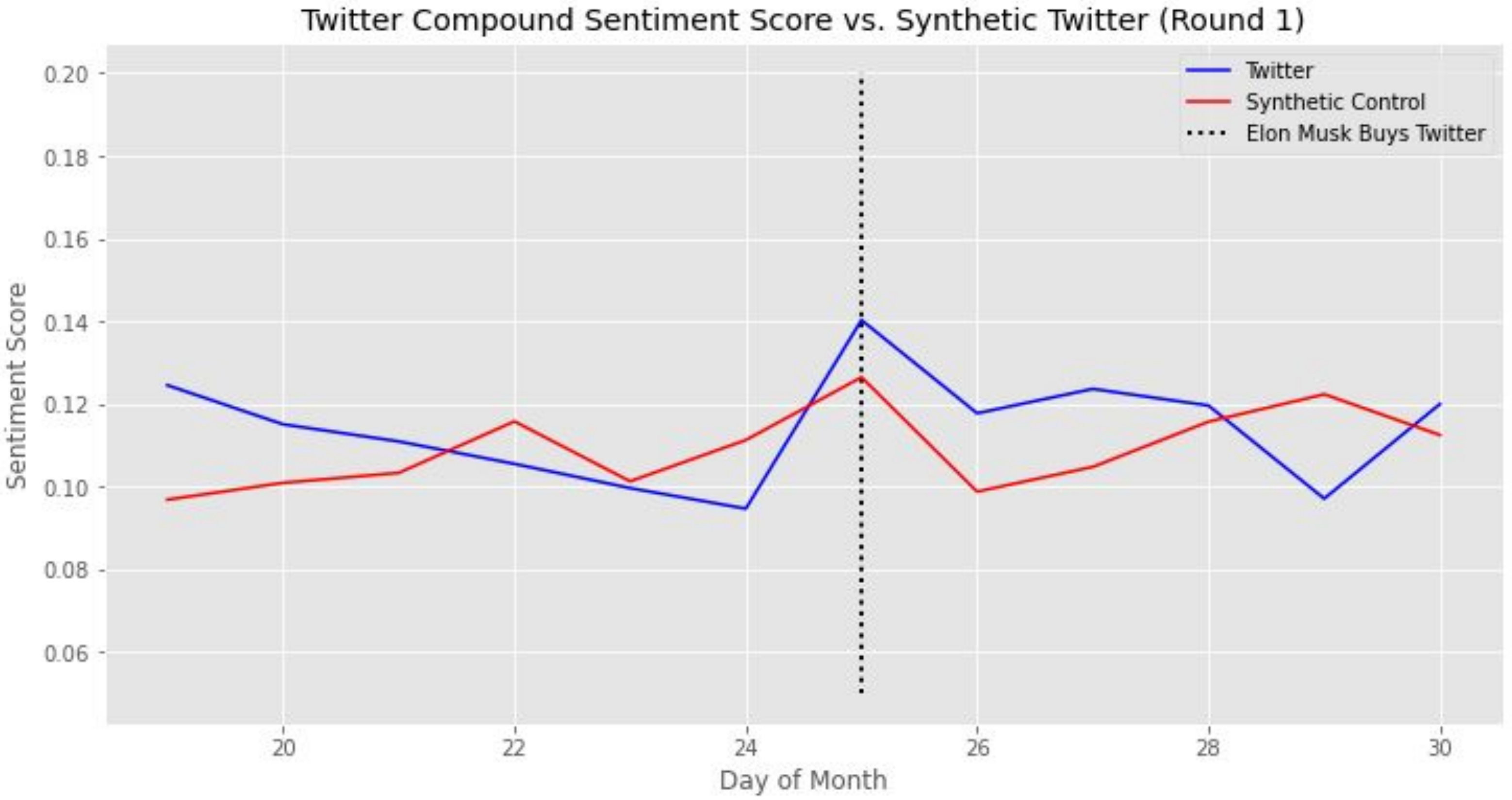
- Importing enough tweets for each company per day to create an accurate Synthetic Twitter
  - At first, we didn’t import enough tweets for an accurate representation
- Getting data into a position where synthetic control and its visualization may be implemented
  - Needed to create a column for each company AKA transpose our original dataframe

Applying Synthetic Control

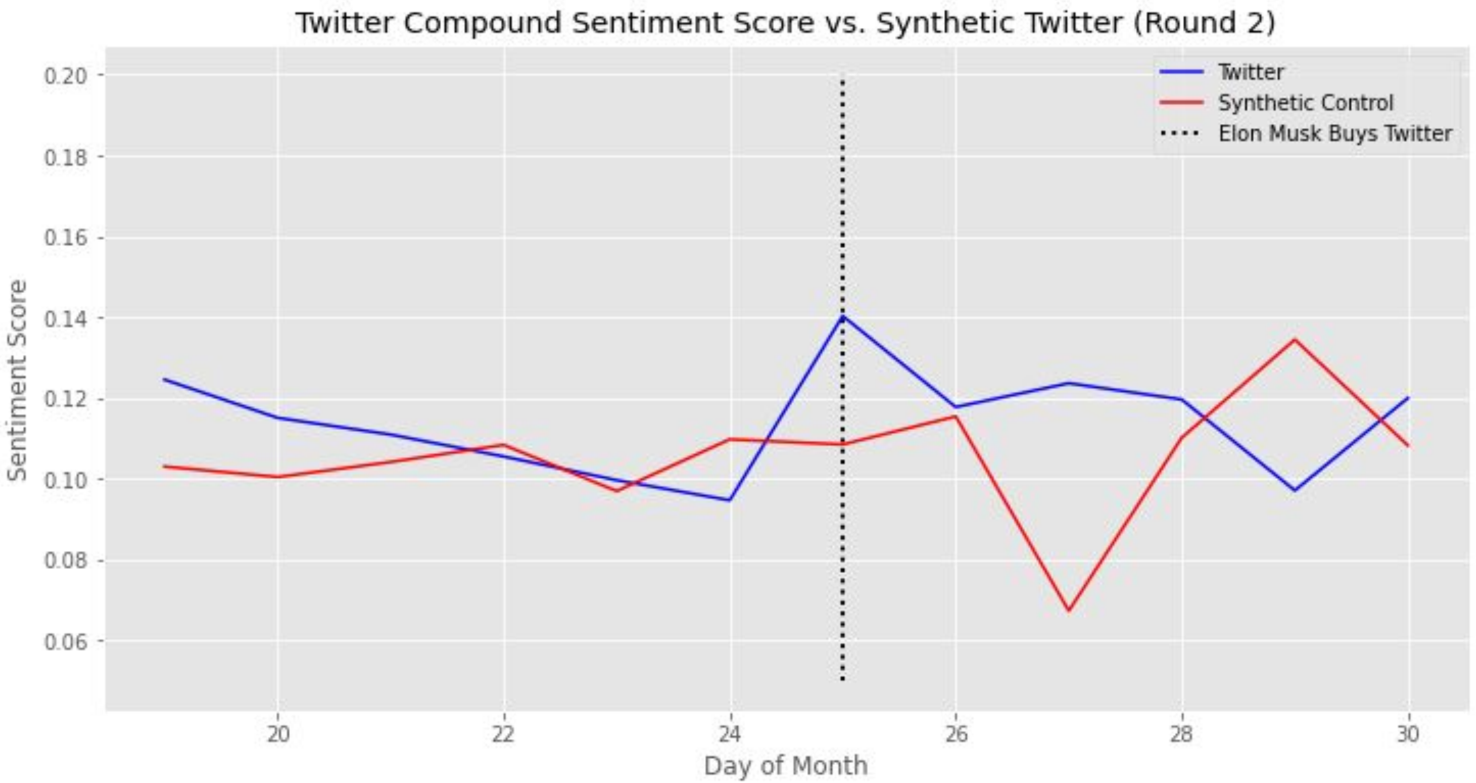
- Construct dataframe like the one below with a column for each company
- Run Linear Regression to calculate weights
- Dot the weights with the specified data to create synthetic Twitter and plot

	company	Facebook	Instagram	LinkedIn	Reddit	Snapchat	Twitter
	Day						
neg	19	0.074993	0.058905	0.053629	0.070246	0.082121	0.077192
	20	0.071859	0.053891	0.044284	0.076901	0.085793	0.078219
	21	0.070761	0.056804	0.048032	0.086847	0.077386	0.080174
	22	0.069179	0.056787	0.046276	0.094155	0.077692	0.081254
	23	0.068549	0.052056	0.047593	0.088407	0.085577	0.086941
	24	0.068891	0.058554	0.041865	0.075534	0.080165	0.081282
pos	19	0.101807	0.105516	0.111208	0.093052	0.104254	0.116656

	Company	Weight
0	Facebook	0.863
1	Instagram	-0.358
2	LinkedIn	0.143
3	Snapchat	0.026
4	Reddit	0.312



Twitter Compound Sentiment Score vs. Synthetic Twitter (Round 1)
Twitter
Synthetic Control
Elon Musk Buys Twitter
Sentiment Score
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
20
22
24
26
28
30
Day of Month



	Company	Weight
0	Facebook	0.2176
1	Instagram	0.0000
2	LinkedIn	0.1798
3	Snapchat	0.0000
4	Reddit	0.6026

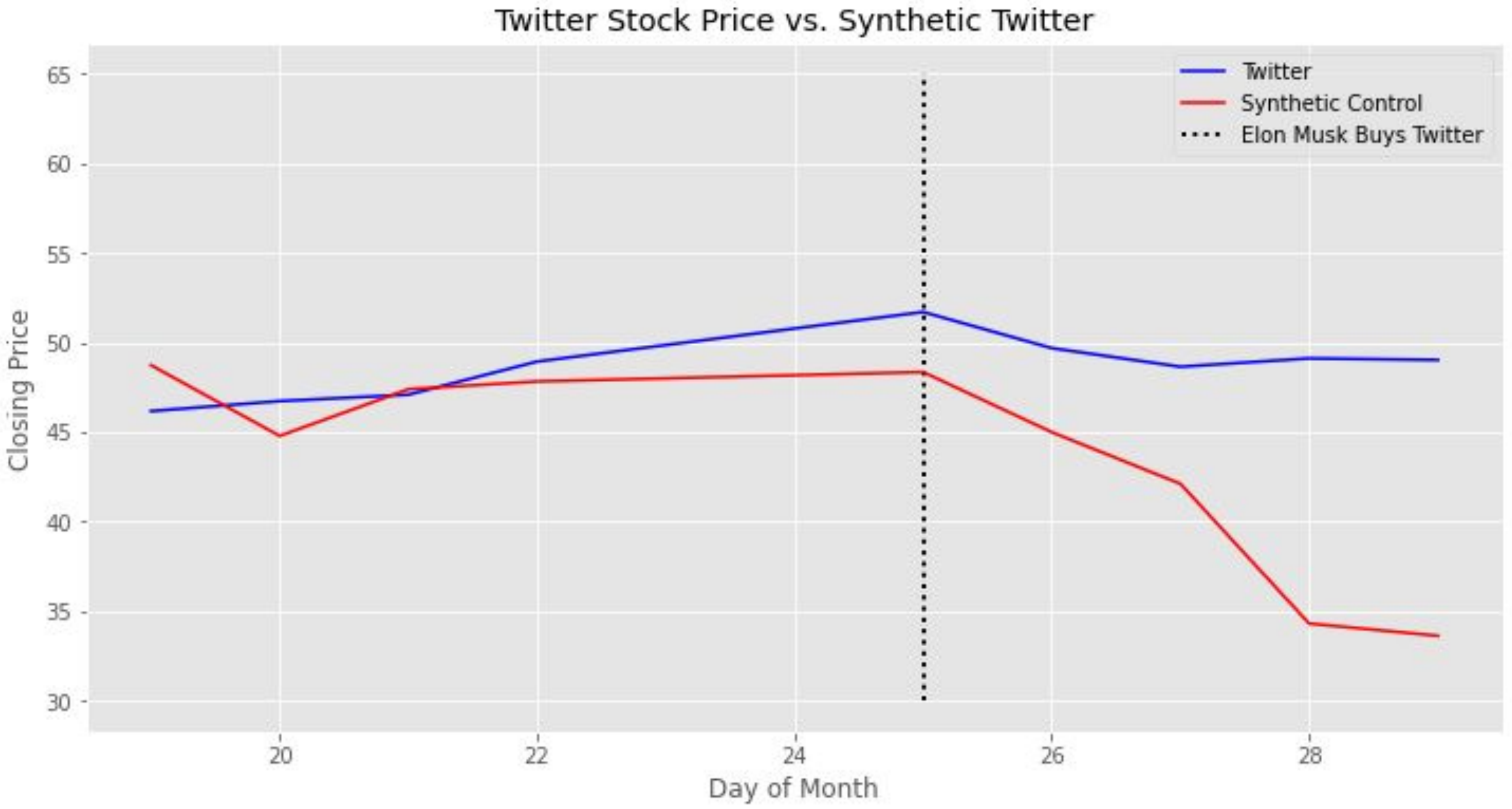
SYNTHETIC CONTROL FOR STOCK PRICE

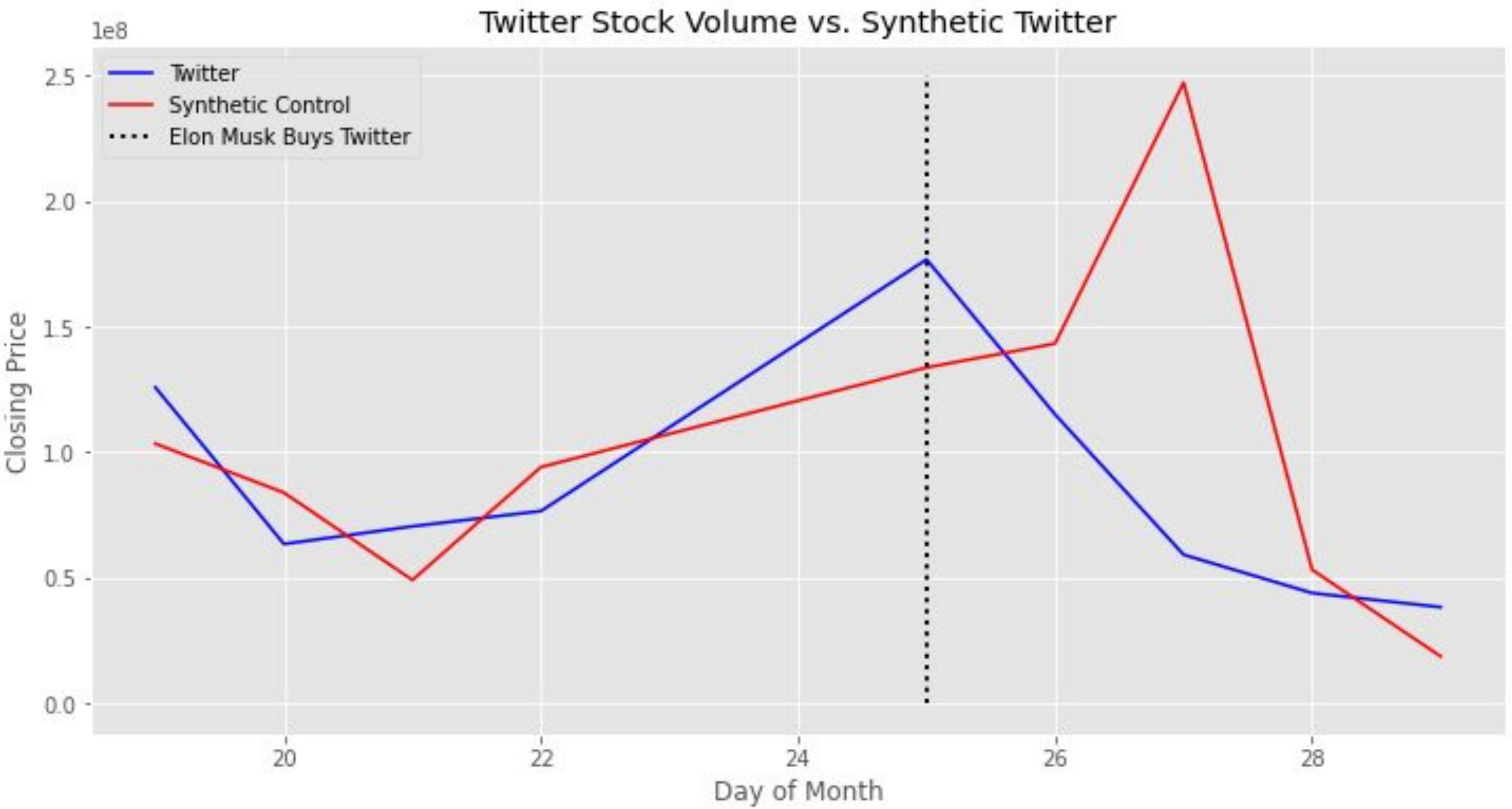
Applying Synthetic Control for Stock Price

- Not all original social media platforms are publicly traded
- We were now limited to Twitter, Facebook, Snapchat, and Pinterest
- Calculated weights again using Linear Regression

	Company	Weight
0	Facebook	-1.041
1	Pinterest	7.439
2	Snapchat	3.152

- Then applied new weights to stock price to create Synthetic Twitter price


Twitter Stock Price vs. Synthetic Twitter
Twitter
Synthetic Control
Elon Musk Buys Twitter
Closing Price
Day of Month
65
60
55
50
45
40
35
30
20
22
24
26
28


Twitter Stock Volume vs. Synthetic Twitter
1e8
2.5
2.0
1.5
1.0
0.5
0.0
Closing Price
20
22
24
26
28
Day of Month
Twitter
Synthetic Control
Elon Musk Buys Twitter

Conclusions

- Applying synthetic control of Twitter Stock Price allowed us to see the treatment effect most clearly.
- Treatment Effect: **Elon Musk purchasing Twitter caused the stock price to increase more than if he did not purchase it.**
- May need better/more “control groups” for effective Sentiment Synthetic
- Compound score was most effective as it is a combination of the other 3 scores, and we could see some treatment effect.
- In the first few days after Musk bought Twitter, we saw a spike in the compound score for Twitter and a decrease in average compound score for the synthetic control, probably surround the buzz from his purchase.