Twitter Analysis With Earning Events

With NLP techniques, the program can analyze the positive and negative sentiments of tweets that contain certain keywords as an indicator of potential price jumps, especially used for pre-earning period. It can also real-time filter the tickers who have strong Twitter volume increase from large ticker pool and send the alert information to multiple email addresses.

Main Functions

Function 1: search raw tweets related to keyword (e.g. TSLA) and store the files by date and analyze the results then visualize the positive or negative sentiment of the tweets with related stock prices, which works as an indicator of potential earning result or price move direction, especially before earning events.

Function 2: get news from specific 30 mainstream news press twitter accounts and store them by date and analyze number of news containing keywords (e.g. COVID) then visualize the result with related prices (e.g. SPY)

Function 3: save ticker names which have earning event in next n days, and sparse the news related to the ticker from seekingalpha.com

Function 4: real-time update the ticker names from specific ticker pool (e.g. S&P 500) which have increasing Twitter volume and their trending detailed information as an indicator of potential large price move, and send the alert information directly to users' email. The tickers names pool is parsed by function3 at each weekend.

Function 5: real-time update the keywords (e.g. stimulus) which have increasing Twitter volume and their trending detailed information as an indicator of potential large price move, and send the alert information directly to users' email.

Function 6: update the keyword topics (e.g. COVID,vaccine) and their trending detailed information as an indicator of potential large price move or price ratio(QQQ/IWM).

Program Structure

The program consists of 4 major modules: processor, statistics, news and visualization; 2 main py files: **menu.py** and **realtime.py**. In each of the modules, several internal python files have their own functions.

📁	data	regular update	2 days ago
📁	dictionary	new function realtime update	9 days ago
📁	examples	transfer news_main to other two main funcitons	last month
📁	main	regular update	2 days ago
📁	news	regular update	2 days ago
📁	processor	regular update	2 days ago
📁	statistics	regular update	15 days ago
📁	visualization	regular update	2 days ago
📄	.gitignore	new function realtime update	9 days ago
📄	LICENSE	Create LICENSE	29 days ago
📄	README.md	updated to v2.1	27 days ago
📄	menu.py	regular update	2 days ago
📄	realtime.py	regular update	2 days ago
📄	run_batch.bat	updated to v2.1	27 days ago

User Instruction and Parameter setting

Function 1,2 and 3 are located in menu.py, function 4 is in realtime.py. To use each function, run the py file with the one function uncommented.

Menu

Function1:

search raw tweets related to keyword (e.g. TSLA) and store the files by date and analyze the results then visualize the positive or negative sentiment of the tweets with related stock prices, which works as an indicator of potential earning result or price move direction, especially before earning events.

parameters:

1. list of tickers

   example: ['AAPL'], the search program will return historical tweets in certain period of time that contain keyword '$AAPL'.

2. dates to lookback

   the maximum date range to search is 8 days for the standard Twitter API account which is being deployed.

3. whether to save the result

   whether or not to save the sentiment result for further study or not.

4. whether to plot

   whether to visualize the result of the sentiment of the analysis result.

5. whether to scale the result with logarithm

   for the high tweets volume result, it is better to visualize the result in a scaled way by applying algorithm to the number which is large.

6. whether to plot stock price with sentiment

   whether or not to plot the sentiment with the intraday data of the stock price, which should be downloaded manually from WRDS TAQ database.

7. whether to parse earning date for each ticker

whether or not to get the most recent earning event date for a ticker.

8. whether to get only the pre-open sentiment

if this parameter is set to 1, the sentiment result time range will only include the time from 4:00PM previous day to 9:30AM of the current date, which is the pre-open time.

9. whether to send the result to certain email address

if this parameter is set to 1, the sentiment analysis result will be sent to certain email addresses.

10. follower threshold

this is the threshold of the sentiment analysis. If the follower of one tweets' author exceed the threshold, the tweet will be counted as one effective tweet. For the less popular companies, the threshold would be set lower to get enough tweets to analyze. The default follower threshold is 5.

11. sentiment analyzer

choose from aggregated classifier model or pure Vader sentiment model. Aggregated classifier model filter the tweets with a pre-filter dictionary; if the sentiment is 0 by the pre-filter dictionary, then the tweet will be analyzed by Vader sentiment tool; if the sentiment is still 0, the tweet will be analyzed by a post-filter dictionary. The later model only use Vader sentiment. Although Vader sentiment model is trained based on social media data, it still omit some sentiment specialized for Twitter and investment content. For example: Vader sentiment would give the sentence 'Big day for $BLK, it is killing now!' a -0.69 basically because of the 'killing' word, but this sentence conveys a positive attitude. 'Our analysis shows Beat for earning of $AAPL. The price will smash the market tomorrow.' is given a 0 sentiment by Vader model, while the pre-filter dictionary will correctly classify it as a strong positive. With the aggregated classifier model, this Twitter specified language pattern being misunderstood could be avoided.

Function2:

get news from specific 30 mainstream news press twitter accounts and store them by date and analyze number of news containing keywords (e.g. COVID) then visualize the result with related prices (e.g. SPY)

parameters:

1. dates to lookback

the maximum date range to search is 8 days for the standard Twitter API account which is being deployed. If searched tweets exceed the range, it will return null result.

2. file name that to be saved

the file name that save all the raw tweets from all the mainstream press Twitter account.

3. keywords need to be studied

the keywords that a tweet contains that would be counted as a effective tweets. For example, if one wants to analyze the relationship of number of coronavirus news and stock price, the keywords that should be set would be 'coronavirus', 'COVID', 'pandemic' and all other words that are related to the coronavirus theme.

Function3:

save ticker names which have earning event in next n days, and sparse the news related to the ticker from seekingalpha.com

parameters:

1. dates to look forward

   N days in the future that one wants to search the companies who have earning events in that future period.

2. index name pool

   ticker names that are limited to a specific name pool. For example the default name pool is S&P500, which means it only search the names in S$P500 list that have earning event in the next N days.

Realtime

Function4:

real-time update the ticker names from specific ticker pool (e.g. S&P 500) which have increasing Twitter volume and their trending detailed information as an indicator of potential large price move, and send the alert information directly to users' email. The tickers names pool is parsed by function3 at each weekend.

parameters:

1. testing timing

   choose what time the monitor function should be run very half hour, the default monitoring time is at each xx:00 and xx:30, for example 8:00AM, 8:30AM, 9:00AM.

2. index name pool

   names that need to be monitored. The default name pool conclude the names that have earning events in the following week.

3. alert trigger condition

   if the trigger condition is reached, the ticker would be recorded and will be added to the alert information. The default trigger condition is when the new half hour volume of one ticker is larger than 2 times the historical half-hour average tweet volume)

4. alert information in email body

   the default alert email body is historical and new volume, positive and negative scores of the historical and new half-hour tweets

5. alert receiver email address

   the default email address is Prof. Rodriguez's Gmail.

Function5:

real-time update the keywords (e.g. stimulus) which have increasing Twitter volume and their trending detailed information as an indicator of potential large price move, and send the alert information directly to users' email.

parameters:

1. monitor topic

   topic need to be monitored. For example, input could be 'Stimulus'. This program will monitor the trend of the tweets that mentioned 'Stimulus' topic.

2. analysis time window

choose how many recent days data we want to analyze

Function6:

update the keyword topics (e.g. COVID,vaccine) and their trending detailed information as an indicator of potential large price move or price ratio(QQQ/IWM).

The analysis depends on the dictionary stored in 'dictionary/MacroTopic.csv' while the twitter accounts where our tweets are from is in the file of 'dictionary/MacroAccounts.csv'. This twitter accounts files have two columns (2020.11.18), the first column contains the twitter accounts of major press in US and worldwide (e.g. ABC news). The second columns are the accounts focusing the COVID topics and also the major press.

By analyzing the recent data from these accounts, we could have an index for each of the topic, representing the mentioned frequency of that topic. By comparing the mention frequency by some price or price ratio(e.g. QQQ/IWM), user can analyze the potential relationship between them to generate future investment signals.

parameters:

1. analysis time window

choose how many recent days data we want to analyze

2. topic dictionary (default)

the dictionary is prerequisite, stored in dictionary folder. The columns are the topic names(e.g. COVID, lockdown). Each column contain the keywords under the topic.
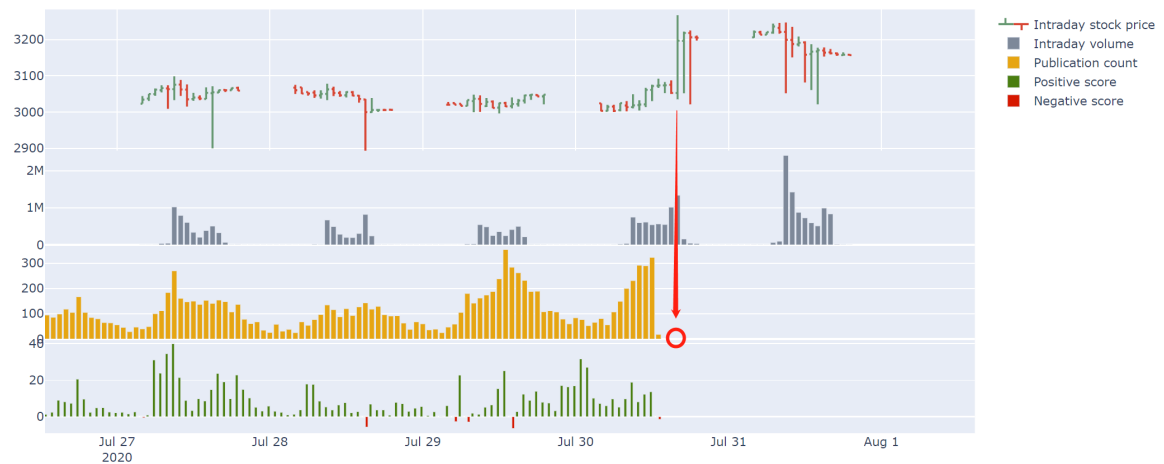
Function Demo

Function1:

Example 1: For AAPL (Apple Inc.), 4:30PM, July 30 is the earning release time. Before the earning event, the overall 24 hours sentiment was very positive with the tweets volume surging at the same time. At the earning release period (4:00PM-5:00PM), the stock price jumped more than 5% after the market trading with a 'Beat' as its earning result .

$AAPL intraday twitter sentiment and earnings info

Example 2: For ticker AMZN (Amazon.com, Inc.), it release the earning event at July 30, 4:30PM. Before the earning event, the overall 24 hours sentiment was very positive with the tweets volume surging at the same time. At the earning release period (4:00PM-5:00PM), the stock price jumped more than 5% after the market trading with a 'Beat' as its earning result .

$AMZN intraday twitter sentiment and earnings info

Example 3: For ticker Z (Zillow Group Inc), the earning release time is after market at 4:30PM, August 6. As is shown in the graph, there was a Twitter volume spike at 11:00AM, which is almost 4 times of the average Twitter volume. With the surge of Twitter volume, the sentiment of the tweets showed positive. At the earning event period (4:00PM to 5:00PM), the stock price jumped more than 10%. The earning event result is also 'Beat'.
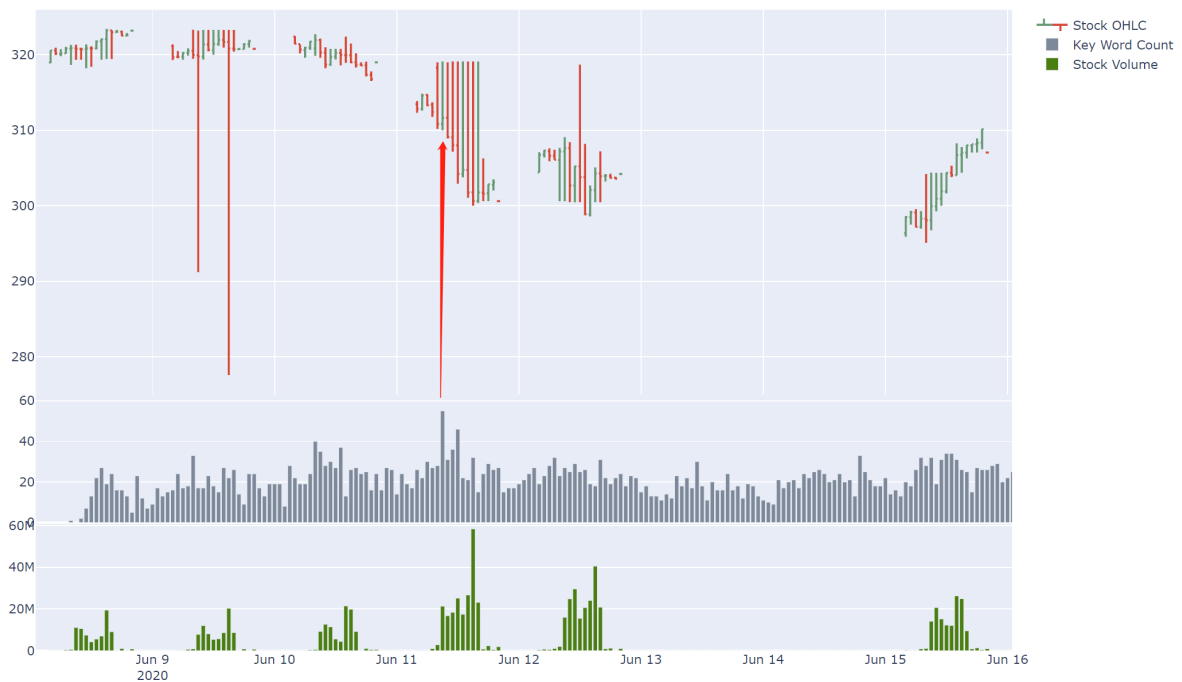

$Z intraday twitter sentiment and earnings info

Function2:

Get all tweets from certain mainstream press accounts including CNN, Fox News, etc. then count tweets number containing certain keywords including Corona, COVID or pandemic, then compare it with SPY intraday graph:

News intraday twitter sentiment

Function3:

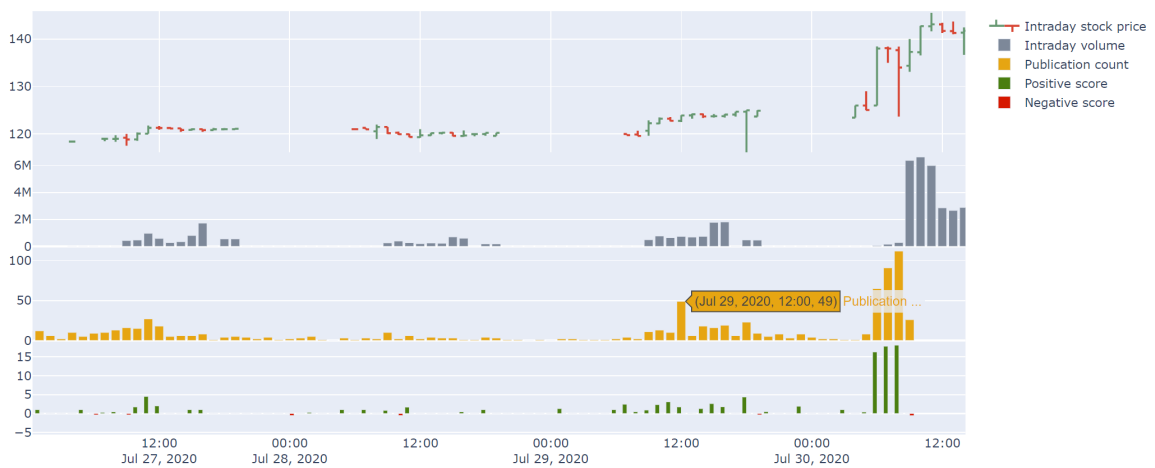The names that have earning events in the next week and the earning release time.

	Name	Date	Time	GICS Sub Industry
MU	Micron Technology, Inc.	2020-06-29	Estimated	Semiconductors
FDX	FedEx Corporation	2020-06-30	Post-Market	Air Freight & Logistics
CAG	Conagra Brands, Inc.	2020-06-30	Pre-Market	Packaged Foods & Meats
STZ	Constellation Brands, Inc.	2020-07-01	Pre-Market	Distillers & Vintners
GIS	General Mills, Inc.	2020-07-01	Estimated	Packaged Foods & Meats
CCL	Carnival Corporation & Plc	2020-07-04	Estimated	Hotels, Resorts & Cruise Lines
PAYX	Paychex, Inc.	2020-07-07	Pre-Market	Data Processing & Outsourced Services
WBA	Walgreens Boots Alliance, Inc.	2020-07-09	Pre-Market	Drug Retail
PEP	PepsiCo, Inc.	2020-07-13	Pre-Market	Soft Drinks
FRC	First Republic Bank	2020-07-13	Estimated	Regional Banks
OMC	Omnicom Group Inc.	2020-07-14	Estimated	Advertising
FAST	Fastenal Company	2020-07-14	Pre-Market	Building Products
WFC	Wells Fargo & Company	2020-07-14	Estimated	Diversified Banks
HWM	Howmet Aerospace Inc.	2020-07-14	Estimated	Aerospace & Defense
C	Citigroup Inc.	2020-07-14	Pre-Market	Diversified Banks
STT	State Street Corporation	2020-07-14	Estimated	Asset Management & Custody Banks
JPM	JPMorgan Chase & Co.	2020-07-14	Pre-Market	Diversified Banks
EBAY	eBay Inc.	2020-07-15	Estimated	Internet & Direct Marketing Retail
ADS	Alliance Data Systems Corporation	2020-07-15	Estimated	Data Processing & Outsourced Services
GS	The Goldman Sachs Group, Inc.	2020-07-15	Pre-Market	Investment Banking & Brokerage
BK	The Bank of New York Mellon Corporation	2020-07-15	Estimated	Asset Management & Custody Banks
MS	Morgan Stanley	2020-07-15	Estimated	Investment Banking & Brokerage
USB	U.S. Bancorp	2020-07-15	Pre-Market	Diversified Banks
PNC	The PNC Financial Services Group, Inc.	2020-07-15	Estimated	Regional Banks
DOV	Dover Corporation	2020-07-15	Estimated	Industrial Machinery
JNJ	Johnson & Johnson	2020-07-16	Estimated	Pharmaceuticals

Function4:

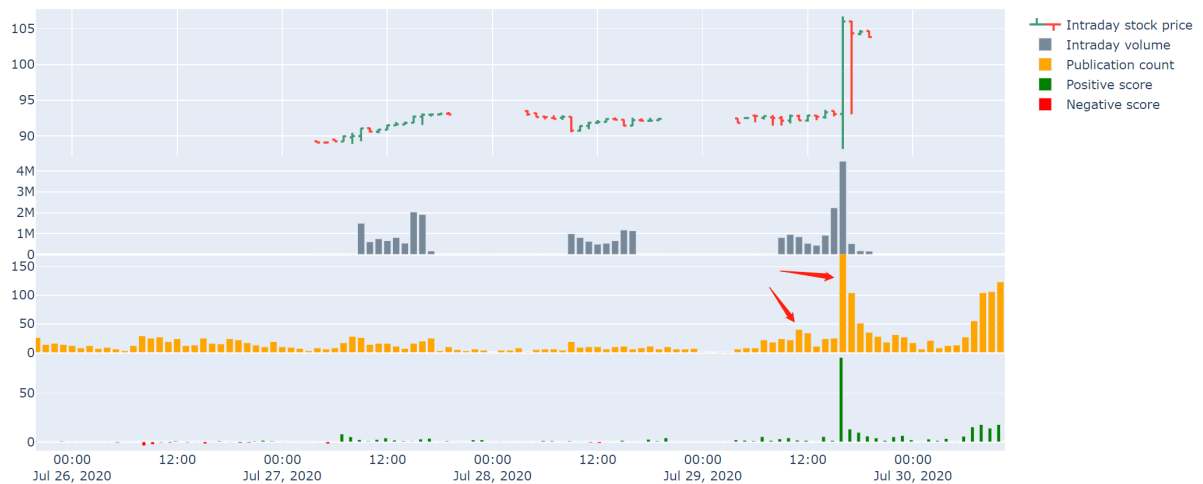The email body contains three ticker names that have high Twitter volume in the past half hour.

Example 1: Before the UPS (United Parcel Service, Inc.) had their earning released before market July 30, there was a surge of Twitter volume between 12:00AM to 1:00PM, which is 5 times compared to the average historical Twitter volume. At the same time, the sentiment of the 24 hour period before the earning event is positive overall. After the earning was released, the stock price jumped up 5% before the market and after the earning event released.

$UPS intraday twitter sentiment and earnings info

Example 2: July 29, after the market, QCOM (QUALCOMM, Inc.) had their earning released and saw a 13% price jump. Before the price jump, there was a 2 times jump of Twitter volume compared to the average historical Twitter volume. The sentiment of the 24 hour period before the earning event is positive overall. This is another example that the alert system could remind the trader about the import event and advice the potential stock move direction.

$QCOM intraday twitter sentiment and earnings info

Example Email body:

[Test] Twitter real time (half) hourly trending alert

noven.zheng@gmail.com   10:05 AM (11 hours ago)
to me, rangerrod1

There are these tickers trending right now: DIS ATVI CVS.   There are the detailed information for the trends:

For $DIS, historical volume: 28.0,   historical positve/negatives score: 18.65/-10.4.   New half hour volume: 242,   positve/negatives score:
23.13/-4.33.

For $ATVI, historical volume: 19.0,   historical positve/negatives score: 20.07/-11.12.   New half hour volume: 43,   positve/negatives score:
24.85/-9.49.

For $CVS, historical volume: 15.0,   historical positve/negatives score: 12.83/-1.74.   New half hour volume: 55,   positve/negatives score:
7.78/-1.82.

Positive vs. Negative tweets

There are some tweets being classified examples:

7/23/2020 7:00	1	127	$SLB gets our "Strong Beat" #EarningsDistortion Score (featured on @SquawkCNBC). Earnings are understated by $14.45/share (>100% of consensus estimates for 2Q20). See other stocks most likely to beat/miss in our Earnings Distortion Scorecard report https://t.co/tj0k9gITym
7/23/2020 7:00	1	2905	$SLB gets our "Strong Beat" #EarningsDistortion Score (featured on @SquawkCNBC). Earnings are understated by $14.45/share (>100% of consensus estimates for 2Q20). See other stocks most likely to beat/miss in our Earnings Distortion Scorecard report https://t.co/O0jOPvUMwq
7/23/2020 7:05	0.4404	411	RT @bi54460186: Discotd https://t.co/VJCTdWfIjw a better, more executive trading chat room $FCX $JBLU $GOLD $LYG $PFE $NLY $VALE $KMI $VIâ€¦
7/23/2020 10:06	1	5705	$SLB earnings tomorrow AM & expected to beat: https://t.co/eh9lgnmAFY Trend fill target: 21.75 https://t.co/IvlumJ6nwV
7/23/2020 12:11	0.5319	141784	$SLB - Schlumberger's Q2 Results Could Get Interesting. https://t.co/hSHMg3lwBT #stocks #finance #economy
7/23/2020 12:17	0.5319	1037	RT @SeekingAlpha: $SLB - Schlumberger's Q2 Results Could Get Interesting. https://t.co/hSHMg3lwBT #stocks #finance #economy
7/23/2020 12:32	1	51	If $SLB beats earnings tomorrow $HAL could see the 200 sma
7/20/2020 2:06	0.4019	6576	RT @StockMarketSam1: It would be interesting to see how these stocks preform after UAE launches interplanetary mission to Mars: $BA $SPEC $â€¦
7/20/2020 2:14	0.7226	21	Market open tomorrow 🤔 Big week for my personal current biggest play $LMT . US Air Force contract broke news after market Friday so hoping to see that play out tomorrow. Still trying to decide how soon Iâ€™m willing to get out tho 🤔 $DOW $NASDAQ #stockstowatch
7/20/2020 5:29	-0.128	1576	Lockheed Martin $LMT PT Lowered to $400.00 at Credit Suisse Group https://t.co/QqkB3NKdca
7/20/2020 5:30	-0.128	1453	Lockheed Martin $LMT PT Lowered to $400.00 https://t.co/LyKxSaSIK9
7/20/2020 5:31	-0.128	2594	Lockheed Martin $LMT PT Lowered to $400.00 at Credit Suisse Group https://t.co/XDZOhGhxrF
7/20/2020 6:10	0.7761	5076	$LMT in Uptrend: price may ascend as a result of having broken its lower Bollinger Band on July 9, 2020. View odds for this and other indicators: https://t.co/V6V2bSmmsh #LockheedMartin #stockmarket #stock #technicalanalysis #money #trading #investing #daytrading #news #today https://t.co/YNJoDUmXY9
7/20/2020 8:35	0.128	266	Expect good things out of $LMT earnings tomorrow. Expectations have been lowered and prices are not that extended.
7/20/2020 9:19	0.1824	12	RT @SeekingAlpha: $LMT - Lockheed Martin: Your Pandemic-Resistant Long-Term Aerospace Play. https://t.co/tBIlioimEE #finance #business #stoâ€¦
7/22/2020 7:01	-0.4019	2903	$TWTR gets our "Strong Miss" #EarningsDistortion Score (featured on @SquawkCNBC). Earnings are overstated by $1.37/share (>100% of consensus estimates for 2Q20). See other stocks most likely to beat/miss in our Earnings Distortion Scorecard report https://t.co/O0jOPvUMwq
7/22/2020 7:01	-0.4019	126	$TWTR gets our "Strong Miss" #EarningsDistortion Score (featured on @SquawkCNBC). Earnings are overstated by $1.37/share (>100% of consensus estimates for 2Q20). See other stocks most likely to beat/miss in our Earnings Distortion Scorecard report https://t.co/tj0k9gITym
7/22/2020 11:24	-0.6486	13	Everyone excited about $TWTR earnings, going long. I will short on that pop when the momentum weakens. Too many lawsuits lurking around, did everyone forget about last week already? And now today...

Function6:

Topic: TOPIC_COVID