

Automotive Companies Stock Volume Context with Tweets and News Articles

Presented by Harrison Kim, Steve Cozine, Christina Nguyen & Darrick Suen

Project Overview

TSLA • NASDAQ



MARKETS
INSIDER



Subscribe

Search markets



U.S. MARKETS OPEN IN 7 HOURS 15 MINUTES

DOW FUTURES

+0.66%

35,445.00

S&P FUTURES

+0.70%

4,622.25

NASDAQ FUTURES

+0.78%

15,967.00

A Jefferies analyst sees Tesla rising 30% in 12 months, even as Elon Musk sells stock. Here's why he has the most bullish call on Wall Street.

Harry Robertson

© Dec. 4, 2021, 05:00 AM



SHARE

+279.24 YTD

+0.74

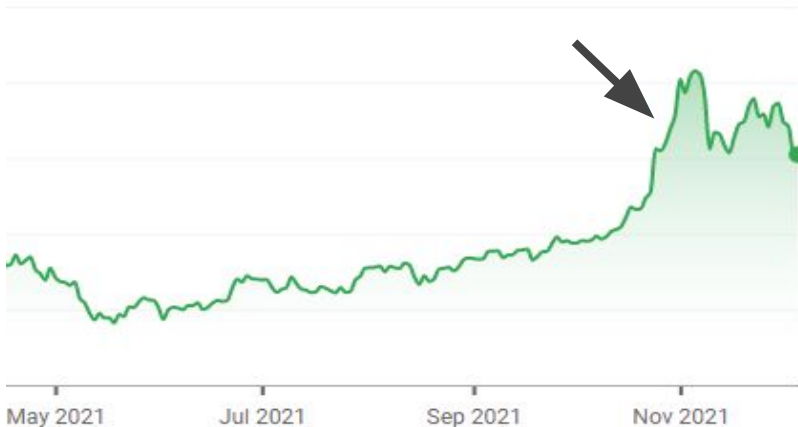
NASDAQ · Disclaimer



1Y

5Y

MAX



ping another
ends that

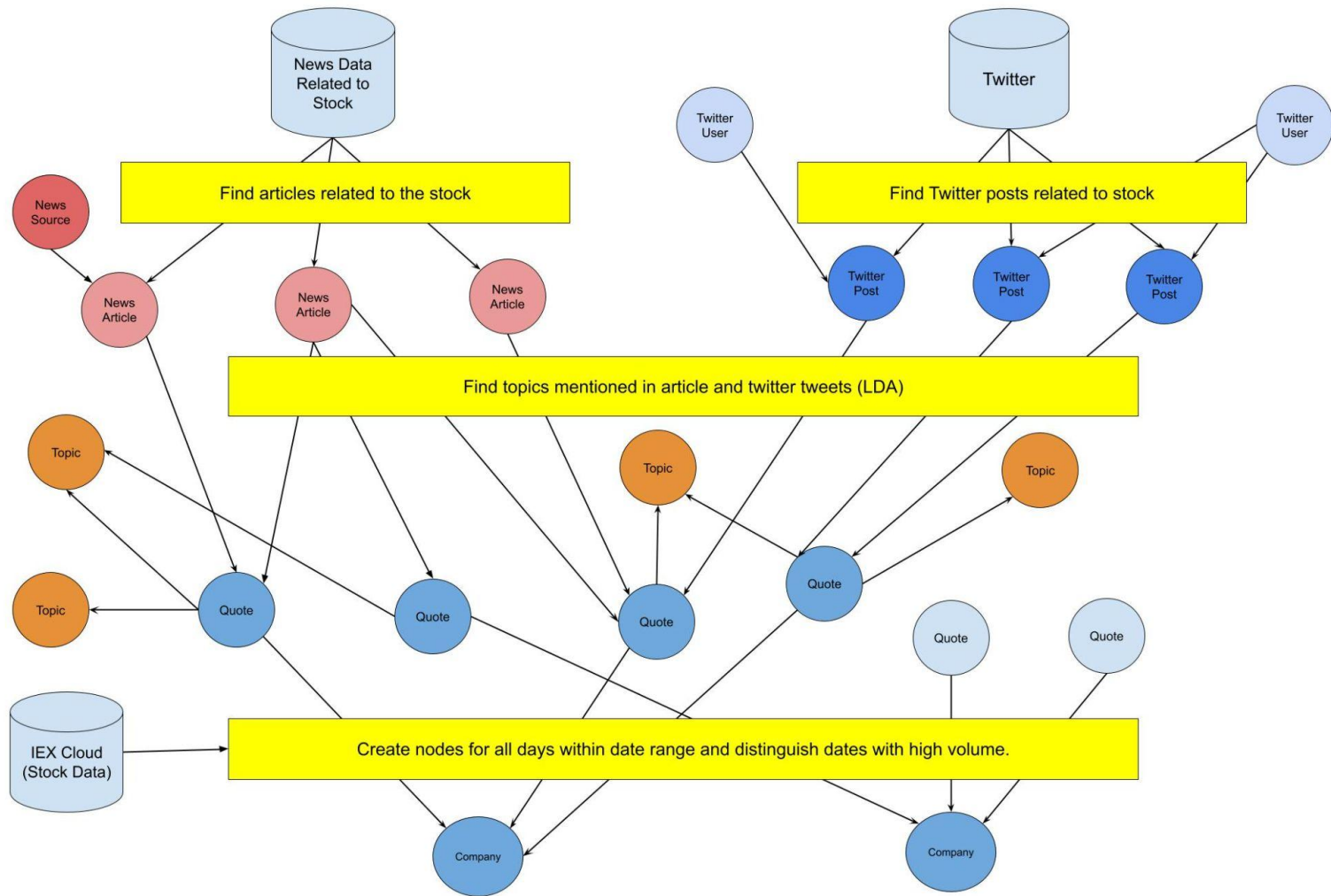
sla could even
ny," in the same

% in 12 months, even...
ergy company' in the
ore,' according to ...

Project Details

- Focused on automotive industry
 - Companies that have more than a year of stock data
 - Companies that are on the US stock market
 - Company list: Ford, Tesla, GM, Stellantis, Honda, Toyota, Ferrari, Tata Motors, NIO
- Focused on these topics:
 - Earnings
 - New product/Product announcement
 - Supply chain
 - Investing analysis
 - Partnership announcement
 - Critique of company performance





Data Sources

Stock Data

- Pull stock data from IEX Cloud API in JSON format
- Structured

HTTP REQUEST

```
GET /stock/{symbol}/company
```

RESPONSE

```
{
  "symbol": "AAPL",
  "companyName": "Apple Inc.",
  "exchange": "NASDAQ",
  "industry": "Telecommunications Equipment",
  "website": "http://www.apple.com",
```

HTTP REQUEST

```
GET /stock/{symbol}/chart/{range}/{date}
```

RESPONSE

```
[
  {
    "close": 116.59,
    "high": 117.49,
    "low": 116.22,
    "open": 116.57,
    "symbol": "AAPL",
    "volume": 46691331,
    "id": "HISTORICAL_PRICES",
    "key": "AAPL",
    "subkey": "",
    "date": "2020-11-27",
```

Tweet Data

- Used package called *snsrape* to pull tweets for the days associated with the stock event
- Semi-structured, but organized tweets into pandas dataframe

	date	tweet_id	content	username	company	company_ticker	lang
0	2021-11-22 22:14:55+00:00	1462907473166225411	nobody:"i don't like scuderia ferrari"\n\n? ht...	vibezwithiaia	Ferrari	\$RACE	en
1	2021-11-22 20:19:23+00:00	1462878398301323279	. @Charles_Leclerc is at Village Paddle Modena...	lovingleclercTV	Ferrari	\$RACE	en
2	2021-11-22 20:27:30+00:00	1462880440071688201	Evil Ferrari pit crew be like: https://t.co/Ey...	JohnF1_	Ferrari	\$RACE	en
3	2021-11-22 22:10:36+00:00	1462906384299663363	This is the moment when Maruti 800 claim to be...	AfaaqKashmiri	Ferrari	\$RACE	en
4	2021-11-22 23:18:07+00:00	1462923375270146060	me every weekend even if Ferrari isn't having ...	f1maxleclerc	Ferrari	\$RACE	en
...
89	2021-11-22 23:14:48+00:00	1462922543522795521	@ScuderiaFerrari @Carlossainz55 @Charles_Lecle...	LeclercFanCL16	Ferrari	\$RACE	en
90	2021-11-22 20:09:36+00:00	1462875935783464970	@ScuderiaFerrari @scuderiarelio @Carlossainz55...	Hurtsvania	Ferrari	\$RACE	en
91	2021-11-22 22:29:26+00:00	1462911127143301125	@F1ClashGame @SchumacherMick @nikita_mazepin @...	nastyniklas	Ferrari	\$RACE	en
92	2021-11-22 20:25:08+00:00	1462879842702176264	@haldin_anton @ScuderiaFerrari @Carlossainz55 ...	bigfanani	Ferrari	\$RACE	en
93	2021-11-22 20:10:56+00:00	1462876270165954560	@sakshuwu @Charles_Leclerc @ScuderiaFerrari th...	CamFerrari25	Ferrari	\$RACE	en

94 rows x 7 columns

News Article Data

- Used *contextual web search* API to pull news articles
 - <https://rapidapi.com/contextualwebsearch/api/web-search>
- Semi-structured
- Extracted: article title, article description, date published, website, URL
- Querying filters:
 - Used list of top 40 news/financial sources like Forbes

The screenshot shows the USA Today website with a dark green header. The main navigation bar includes links for News, Sports, Entertainment, Life, Money (highlighted), Tech, Travel, and Opinion. At the top, there are links for 'TAKE OUR RETIREMENT SAVINGS QUIZ', 'Retirement basics', 'YOUNG INVESTORS', 'Crypto risk, reward', and 'INVESTIGATION'. The article is titled 'GM to make Chevrolet Silverado electric pickup, competing with electric Ford F-150' under the 'CARS' category. The author is Jamie L. LaReau from the Detroit Free Press, and it was published on April 6, 2021. A 'Node Properties' sidebar on the right lists article details in a table.

Node Properties	
<id>	17679
date_published	2021-04-06
description	GM announced it will offer an all-electric Chevrolet pickup, which will be built at its factory in Detroit and Hamtramck, Michigan.
id	3592195180879507557
language	en
title	Chevrolet Silverado goes electric: GM to make EV version of pickup
url	http://rssfeeds.usatoday.com/~648554524/0/usatoday-techtopstories-GM-to-make-Chevrolet-Silverado-electric-pickup-competing-with-electric-Ford-F/

Creation of Neo4j database

User Input

(Ticker, Start Date, End Date)

```
if __name__ == "__main__":  
    app = App(uri, user, password)  
    app.create_stock('F', dt.datetime(month = 1, day = 1, year = 2021), dt.datetime(month = 11, day = 30, year = 2021))  
    app.create_stock('TSLA', dt.datetime(month = 1, day = 1, year = 2021), dt.datetime(month = 11, day = 30, year = 2021))  
    app.create_stock('GM', dt.datetime(month = 1, day = 1, year = 2021), dt.datetime(month = 11, day = 30, year = 2021))  
    app.create_stock('STLA', dt.datetime(month = 1, day = 1, year = 2021), dt.datetime(month = 11, day = 30, year = 2021))
```

Node Creation Process

Stock Node -> Quote Node -> Article Node -> New Source Node -> Topic Node
-> Tweet Node -> User Node -> Topic Node

Stock & Quote Nodes

Stock Node

Properties





- {compName, industry} - From IEX API given stock ticker
- {ticker} - User Input

Quote Node








Properties

- {close, date, event, high, low, open, volume} - from IEX API for every day in date range
- {day} - Weekday that quote for stock occurred
- {event} - True if volume of the node exceeds 2x the average of volume of the same weekday and stock, otherwise false
- {ticker} - User Input

Stock

<id>	6367	
compName	Ford Motor Co.	
industry	Automobile Manufacturing	
ticker	F	

Quote

<id>	7135	
close	12.49	
date	2021-03-16	
day	Tuesday	
event	False	
high	13.15	
low	12.43	
open	13.06	
ticker	F	
volume	125425273	

Article and News Source Nodes

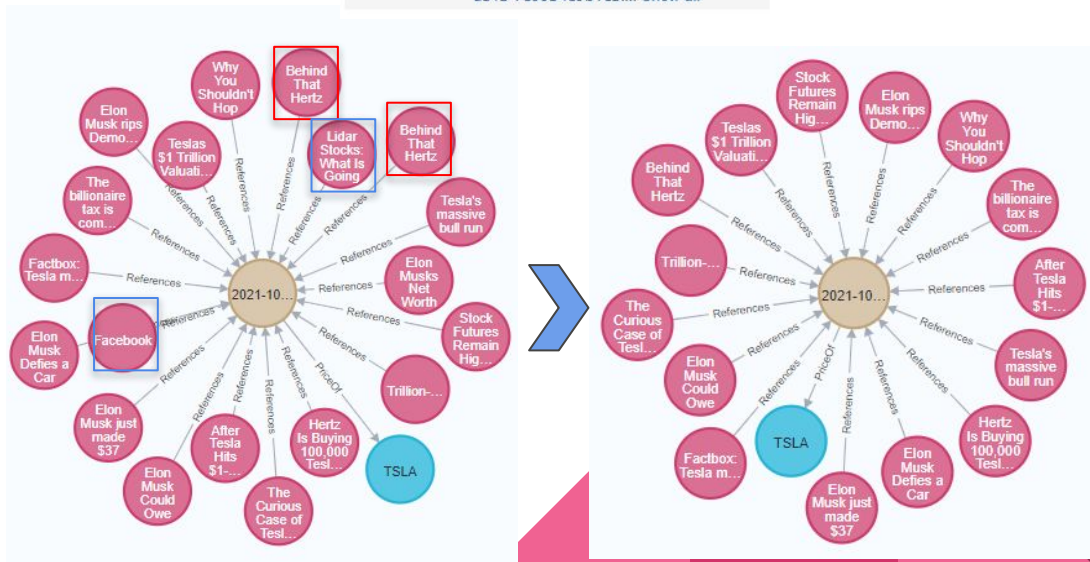
Properties

- {date_published, description, id, language, title, url} - From Article API

Transformation

- Duplicate article removal
- Removal of non ticker specific articles

Article		
<id>	13608	
date_published	2021-04-06	
description	The Detroit theatre of the electric mobility war just got more interesting. General Motors has announced that it will build the all-new full-size Chev... Show all	
id	4823220049414716609	
title	GM to Build Chevy Silverado Electric Pickup at Factory Zero	
url	http://wfmz.com/online_features/green_living/gm-to-build-chevy-silverado-electric-pickup-at-factory-zero/article_a61d12db-fe84-5d4c-a513-789934e9b1e3.... Show all	



Post and User Nodes

Properties

- {date_published, description, id, language, title, url} - From Snsrape API, using comp name

Transformation

- Duplicate tweet removal

"Twitter and General Motors eyes #bitcoin, may expose their portfolio to the digital gold By BTC Peers #gold <https://t.co/ISUtnBuV67> <https://t.co/sPcf0als2c>" (1359610650893447173)





"Twitter and General Motors eyes Bitcoin, may expose their portfolio to the digital #gold - <https://t.co/CQN2aJw90y>" (1359610621512327174)

"About 44,000 members of the United Auto Workers union will be getting \$9,000 checks as a part of a profit-sharing agreement after General Motors brought in \$9 billion in revenue in North America. <https://t.co/VnWJpjXe5f>" (1359612919802191874)

"About 44,000 members of the United Auto Workers union will be getting \$9,000 checks as a part of a profit-sharing agreement after General Motors brought in \$9 billion in revenue in North America. <https://t.co/WmQluCwkyS>" (1359635894056157184)

"About 44,000 members of the United Auto Workers union will be getting \$9,000 checks as a part of a profit-sharing agreement after General Motors brought in \$9 billion in revenue in North America. <https://t.co/tCHCb0e9YS>" (1359622053473648642)

Post

<id>	4189	
content	Cruise, the self-driving startup company backed by General Motors, is getting help from a tech titan as it tries to speed up the commercialization of ... Show all	
date	2021-01-20	
tweet_id	1351806670360489984	

Topic Node

Properties

- {name}

Transformation

- Topic Modelling using LDA
- Remove links and @Twitterhandles from Tweets

Topic

<id>	7317	
name	supply chain	

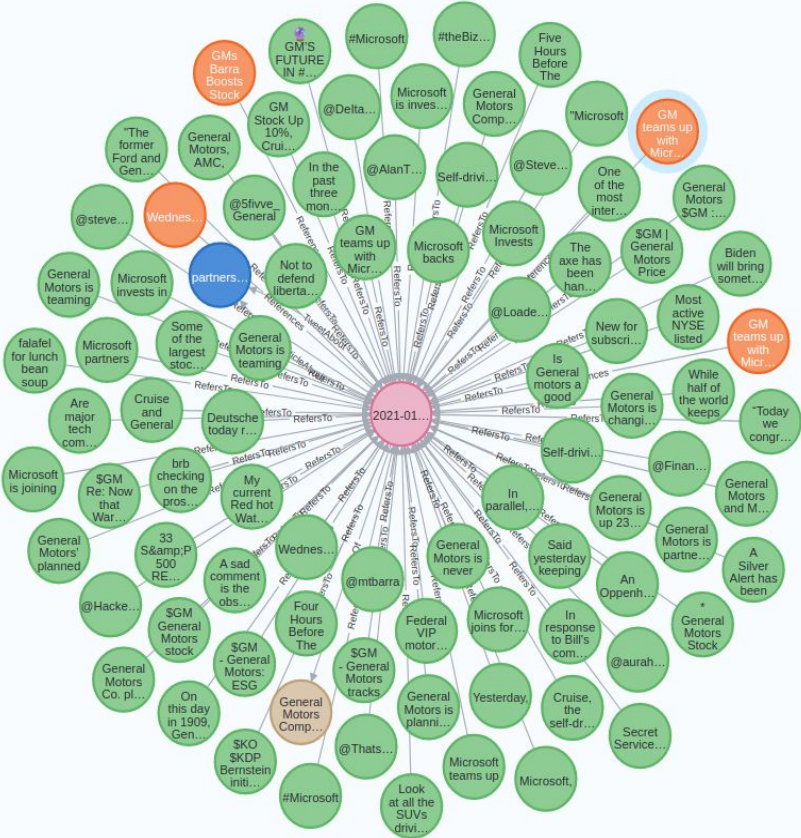
- Conceptually, if we perform LDA on a set of Tweets/Articles from one day, we should be able to find the “topic” of that day
- We apply this methodology to our stock event days
- With a set of day-of-tweets we clean, tokenize, and lemmatize each tweet and combine all into a single list, then process into distinct topics using LDA (we generated 3 per day-of-Tweets)
- Manually label a number of these topics (i.e. “Global”, “Chip”, “Semiconductor”, “Shortage” into “Supply Chain Topic”)
- Combine topics of the same type by doing a full outer join on words, and combining the weights
- Use these master sets to evaluate unlabelled Tweets from an event day with an inner join and combining weights
- The max of the combined weights is the assigned topic
- Repeat for article descriptions

Node Properties

Post		
<id>	4317	
content	GM teams up with Microsoft as electric car market revs General Motors announced an alliance Tuesday with Microsoft on its Cruise autonomous driving ve... Show all	
date	2021-01-20	
tweet_id	1351937565797048321	

Node Properties

Post		
<id>	4210	
content	Microsoft is joining Cruise, the self-driving subsidiary of General Motors, to help speed up the commercialization of autonomous vehicle with a new eq... Show all	
date	2021-01-20	
tweet_id	1351726345395478529	



Node Properties

Article		
<id>	4169	
date_published	2021-01-20	
description	General Motors is teaming up with Microsoft to accelerate its rollout of electric, self-driving cars. In the partnership announced Tuesday, the compan... Show all	
id	3040156220434616270	
title	GM teams up with Microsoft on driverless cars	
url	https://www.onenewspage.com/n/Technology/1zmzz1l2c6/GM-teams-up-with-Microsoft-on-driverless-cars.htm	

Node Properties

Article		
<id>	4172	
date_published	2021-01-20	
description	San Francisco (AFP) Jan 19, 2021 General Motors announced an alliance Tuesday with Microsoft on its Cruise autonomous driving venture, combining forc... Show all	
id	111046974055625975	
title	GM teams up with Microsoft as electric car market revs	
url	https://www.onenewspage.com/n/Science/1zmzz1klap/GM-teams-up-with-Microsoft-as-electric-car.htm	

Demonstration of Graph and Report

Neo4j Queries

- Search for specific Stock, Events, and Articles/Tweets

```
MATCH (s:Stock {ticker: 'F'})-[]-(e:Quote)-[]-(n) WHERE e.event = 'True' RETURN *
```

- Search for Articles or Topics that Reference More than One Event Node

```
MATCH (s:Stock {ticker: 'F'})-[]-(e:Quote)-[]-(a:Article) WHERE e.event = 'True' RETURN *
```

```
MATCH (s:Stock {ticker: 'F'})-[]-(e:Quote)-[]-(t:Topic) WHERE e.event = 'True' RETURN *
```

- Articles that Referenced Two or More Stocks

```
MATCH (s1:Stock)-[]-(q1:Quote)-[]-(a:Article)-[]-(q2:Quote)-[]-(s2:Stock) RETURN *
```


- Stocks with Topics in Common

```
MATCH (s1:Stock)-[]-(q1:Quote)-[]-(t:Topic)-[]-(q2:Quote)-[]-(s2:Stock) RETURN *
```

- All Articles from a Specific Source

```
MATCH (so:Source{name:"reuters"})-[]-(a:Article)-[]-(n)-[]-(s:Stock) RETURN *
```

Lessons Learned & Future Work

- Introduction to determining topics from LDA
 - How to prototype graph schemas before implementation
 - Creating a neo4j graph with tweets, articles, and connected with stock events
 - Addressing challenges with filtering text data from different sources
-
- Automate model training
 - Pulling more years of stock
 - Improving topic model
 - Implement further parallelization to speed up API requests, graph creation
 - Try on different industries
 - Drawing more conclusions on analysis
- 



Q&A