Extracting Trending Financial News From Daily News Articles

SOURAV DEY (IIT2020159)

Introduction

In this presentation, we will explore the techniques and tools used for finding trending news articles, and the benefits of using these techniques in practice. We will also provide real-world examples to illustrate how these techniques work in practice, and how they can be used to monitor media coverage of specific topics, stay up-to-date on the latest news, and identify emerging trends.

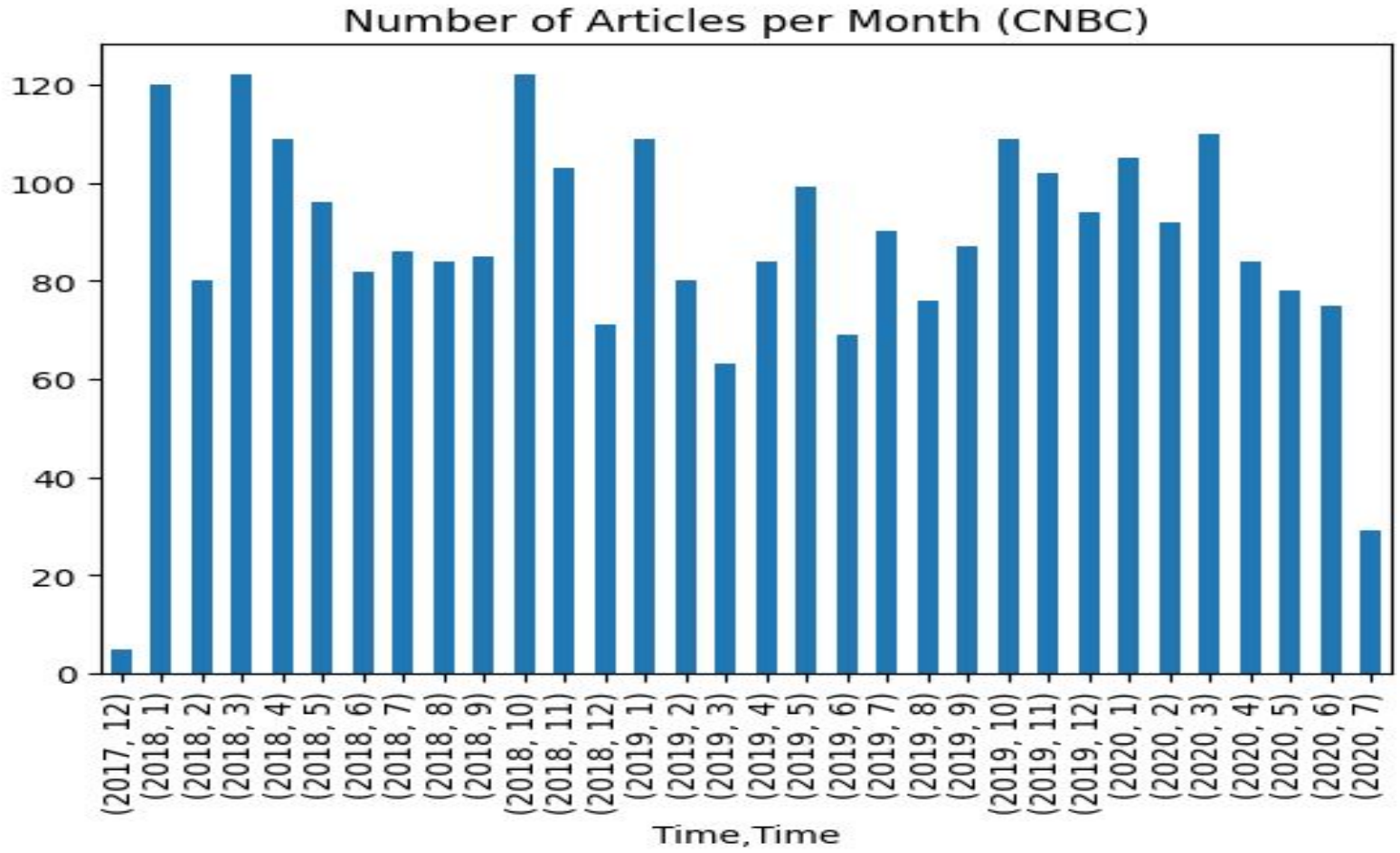
Dataset

The Dataset is scraped from CNBC, the Guardian, and Reuter's official websites, the headlines in these datasets reflects the overview of the U.S. economy and stock market every day for the past year to 2 years.

Columns Provided in the Dataset :-

a) time
b) headlines
c) Description

Analysis Of The Dataset


Number of Articles per Month (CNBC)
120
100
80
60
40
20
0
(2017, 12)
(2018, 1)
(2018, 2)
(2018, 3)
(2018, 4)
(2018, 5)
(2018, 6)
(2018, 7)
(2018, 8)
(2018, 9)
(2018, 10)
(2018, 11)
(2018, 12)
(2019, 1)
(2019, 2)
(2019, 3)
(2019, 4)
(2019, 5)
(2019, 6)
(2019, 7)
(2019, 8)
(2019, 9)
(2019, 10)
(2019, 11)
(2019, 12)
(2020, 1)
(2020, 2)
(2020, 3)
(2020, 4)
(2020, 5)
(2020, 6)
(2020, 7)
Time,Time

Frequent KeyWords using Wordcloud

Methodology

Data Collection: Gather a large collection of news articles from various sources, such as news websites, blogs, and social media platforms. Use web scraping tools to automate the process of collecting articles.

Data Preprocessing: Clean the data by removing stop words, punctuations, and special characters. Also, tokenize the articles by splitting them into words or phrases, and then convert the text to lowercase.

Keyword Extraction: Apply keyword extraction techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) or TextRank or YAKE to identify the most relevant keywords or phrases in each article. You can use these keywords to group similar articles together and identify trending topics.

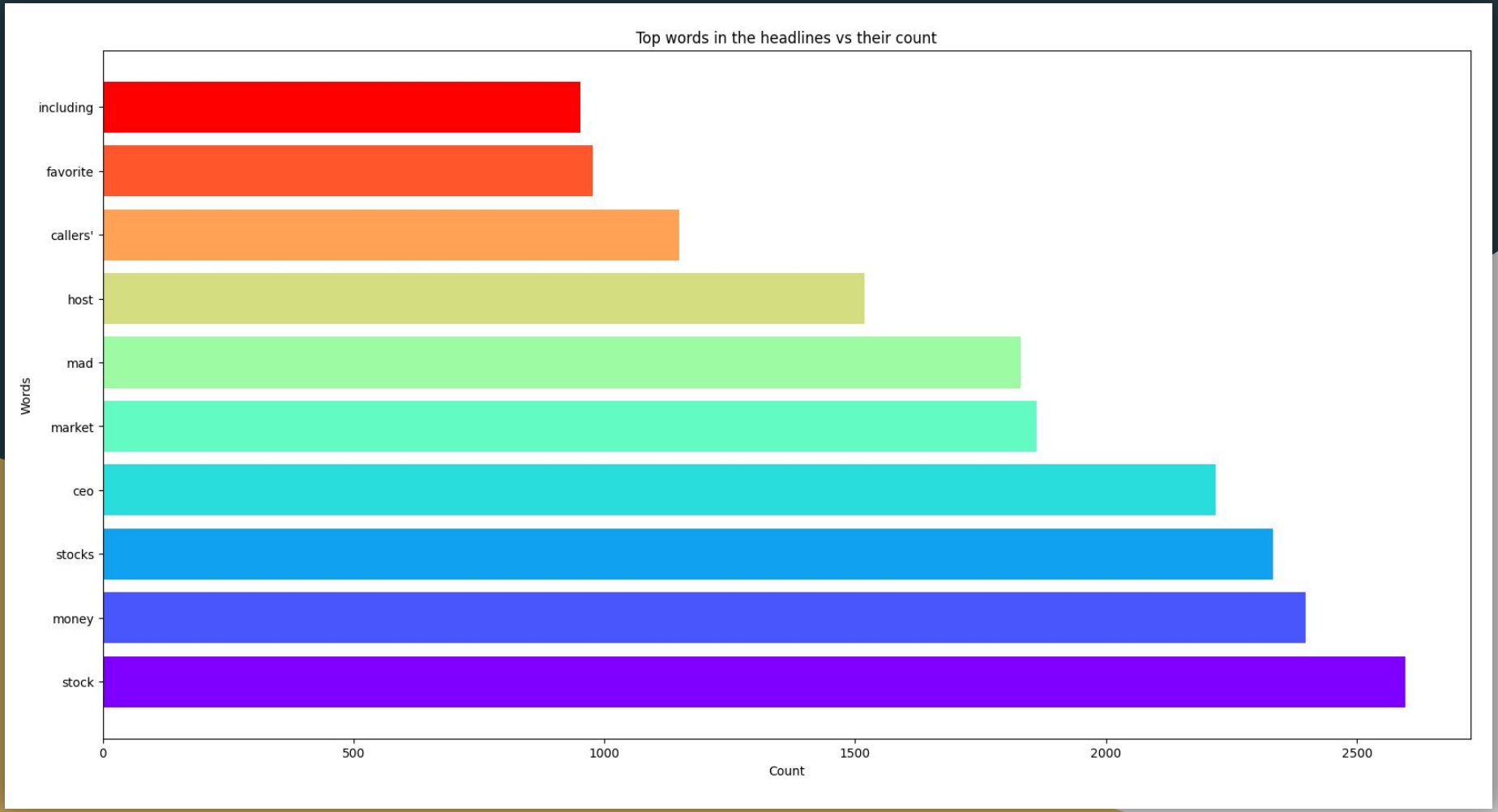
Clustering: Use unsupervised learning algorithms such as k-means, hierarchical clustering, or DBscan to cluster the news articles based on their similarity. These algorithms group similar articles together, making it easier to identify trending topics.

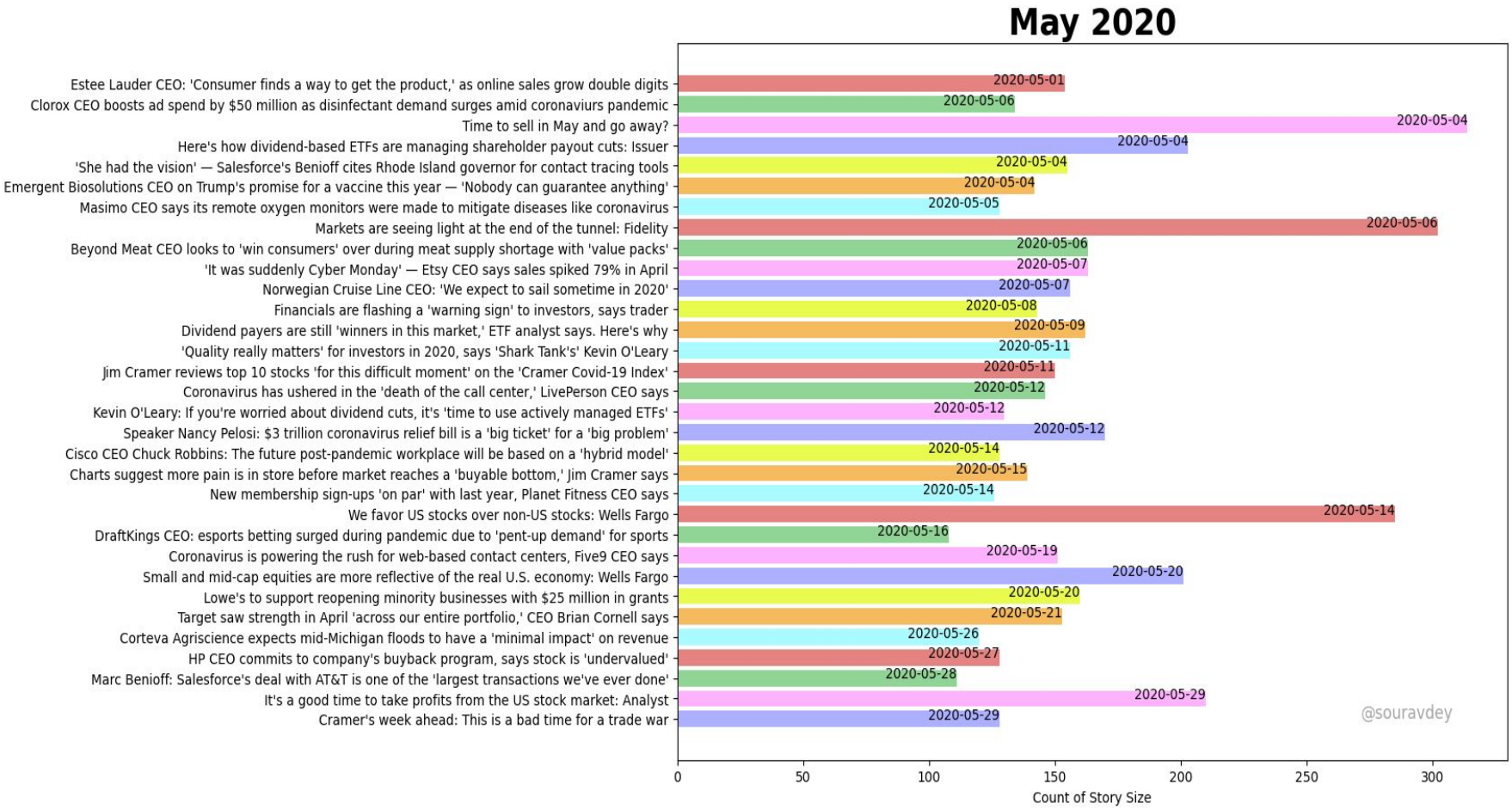
Analysis: Examine the clusters of news articles to identify the most frequent and important topics. You can also analyze the sentiment of the articles to determine whether the stories are positive, negative, or neutral.

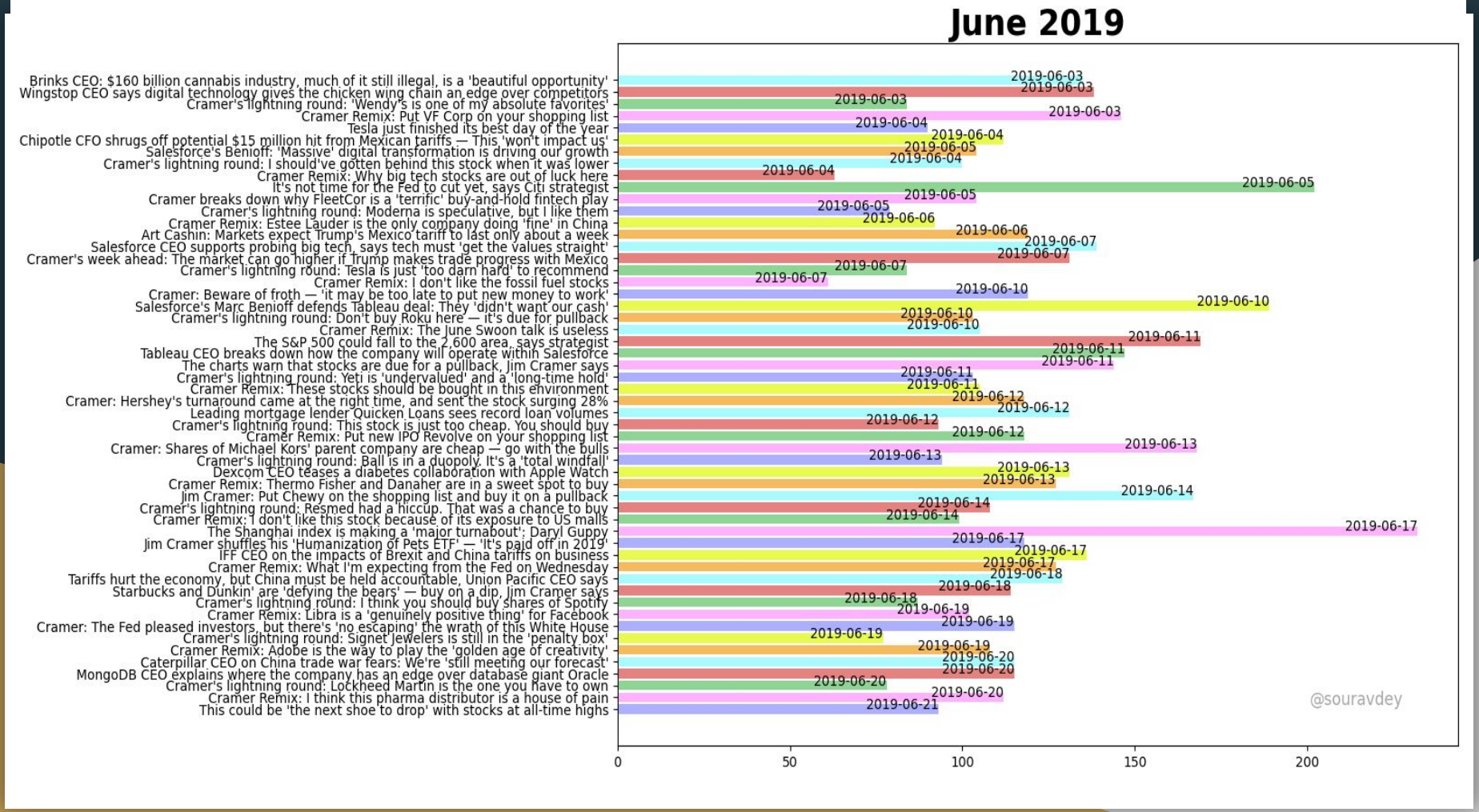
Visualization: Use data visualization tools such as word clouds, scatterplots, and network graphs to display the results in an easy-to-understand format. This can help you identify patterns and trends in the data more easily.

Results

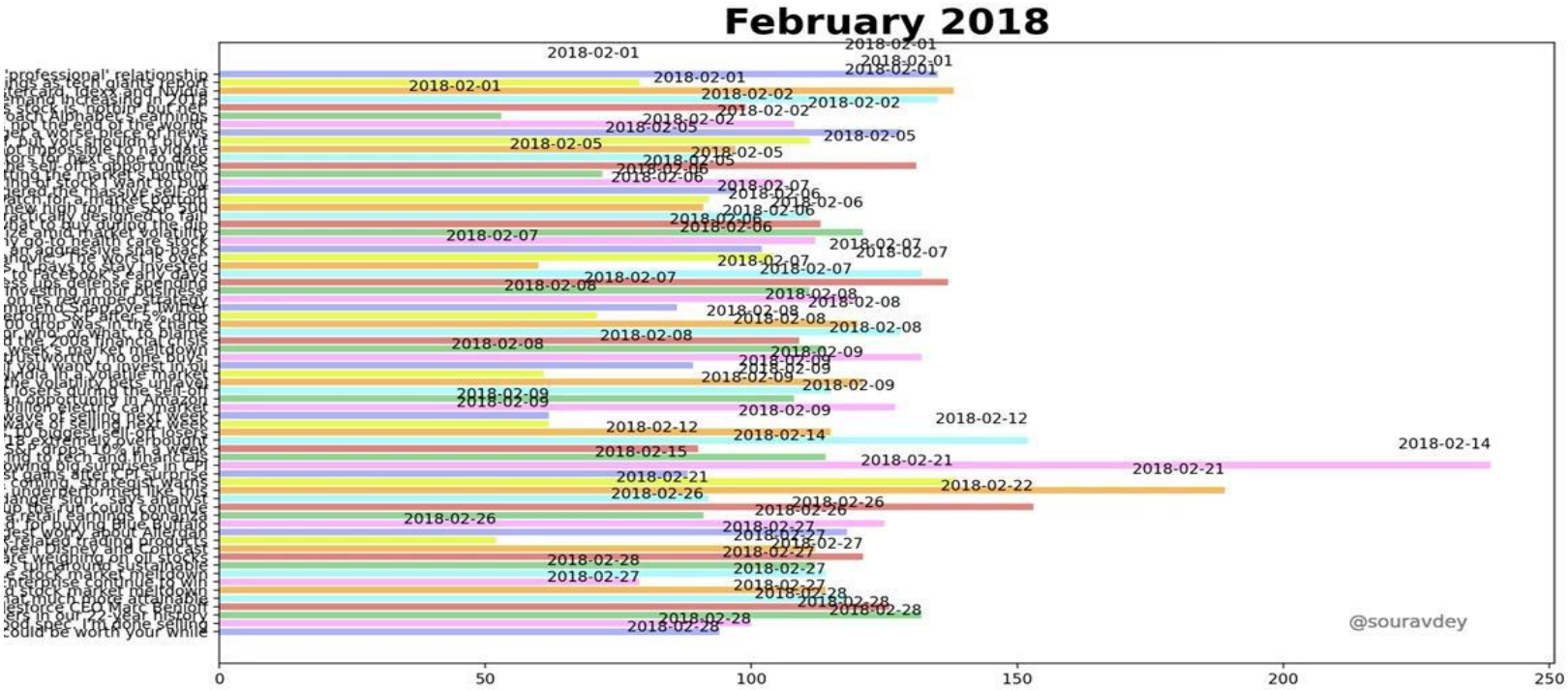
After news articles are clustered, the trending stories are extracted straightforwardly. For any given date, the trending stories here are defined as the top clusters with more news articles published in the past certain time period, e.g., the past two weeks.


Top words in the headlines vs their count
including
favorite
callers'
host
mad
market
ceo
stocks
money
stock
Words
0
500
1000
1500
2000
2500
Count


May 2020
Estee Lauder CEO: 'Consumer finds a way to get the product,' as online sales grow double digits
Clorox CEO boosts ad spend by $50 million as disinfectant demand surges amid coronaviurs pandemic
Time to sell in May and go away?
Here's how dividend-based ETFs are managing shareholder payout cuts: Issuer
'She had the vision' — Salesforce's Benioff cites Rhode Island governor for contact tracing tools
Emergent Biosolutions CEO on Trump's promise for a vaccine this year — 'Nobody can guarantee anything'
Masimo CEO says its remote oxygen monitors were made to mitigate diseases like coronavirus
Markets are seeing light at the end of the tunnel: Fidelity
Beyond Meat CEO looks to 'win consumers' over during meat supply shortage with 'value packs'
'It was suddenly Cyber Monday' — Etsy CEO says sales spiked 79% in April
Norwegian Cruise Line CEO: 'We expect to sail sometime in 2020'
Financials are flashing a 'warning sign' to investors, says trader
Dividend payers are still 'winners in this market,' ETF analyst says. Here's why
'Quality really matters' for investors in 2020, says 'Shark Tank's' Kevin O'Leary
Jim Cramer reviews top 10 stocks 'for this difficult moment' on the 'Cramer Covid-19 Index'
Coronavirus has ushered in the 'death of the call center,' LivePerson CEO says
Kevin O'Leary: If you're worried about dividend cuts, it's 'time to use actively managed ETFs'
Speaker Nancy Pelosi: $3 trillion coronavirus relief bill is a 'big ticket' for a 'big problem'
Cisco CEO Chuck Robbins: The future post-pandemic workplace will be based on a 'hybrid model'
Charts suggest more pain is in store before market reaches a 'buyable bottom,' Jim Cramer says
New membership sign-ups 'on par' with last year, Planet Fitness CEO says
We favor US stocks over non-US stocks: Wells Fargo
DraftKings CEO: esports betting surged during pandemic due to 'pent-up demand' for sports
Coronavirus is powering the rush for web-based contact centers, Five9 CEO says
Small and mid-cap equities are more reflective of the real U.S. economy: Wells Fargo
Lowe's to support reopening minority businesses with $25 million in grants
Target saw strength in April 'across our entire portfolio,' CEO Brian Cornell says
Corteva Agriscience expects mid-Michigan floods to have a 'minimal impact' on revenue
HP CEO commits to company's buyback program, says stock is 'undervalued'
Marc Benioff: Salesforce's deal with AT&T is one of the 'largest transactions we've ever done'
It's a good time to take profits from the US stock market: Analyst
Cramer's week ahead: This is a bad time for a trade war
2020-05-01
2020-05-06
2020-05-04
2020-05-04
2020-05-04
2020-05-04
2020-05-05
2020-05-06
2020-05-06
2020-05-07
2020-05-07
2020-05-08
2020-05-09
2020-05-11
2020-05-11
2020-05-12
2020-05-12
2020-05-12
2020-05-14
2020-05-15
2020-05-14
2020-05-14
2020-05-16
2020-05-19
2020-05-20
2020-05-20
2020-05-21
2020-05-26
2020-05-27
2020-05-28
2020-05-29
2020-05-29
@souravdey
0
50
100
150
200
250
300
Count of Story Size


June 2019
Brinks CEO: $160 billion cannabis industry, much of it still illegal, is a 'beautiful opportunity'
Wingstop CEO says digital technology gives the chicken wing chain an edge over competitors
Cramer's lightning round: 'Wendy's is one of my absolute favorites'
Cramer Remix: Put VF Corp on your shopping list
Tesla just finished its best day of the year
Chipotle CFO shrugs off potential $15 million hit from Mexican tariffs — This 'won't impact us'
Salesforce's Benioff: 'Massive' digital transformation is driving our growth
Cramer's lightning round: I should've gotten behind this stock when it was lower
Cramer Remix: Why big tech stocks are out of luck here
It's not time for the Fed to cut yet, says Citi strategist
Cramer breaks down why FleetCor is a 'terrific' buy-and-hold fintech play
Cramer's lightning round: Moderna is speculative, but I like them
Cramer Remix: Estee Lauder is the only company doing 'fine' in China
Art Cashin: Markets expect Trump's Mexico tariff to last only about a week
Salesforce CEO supports probing big tech, says tech must 'get the values straight'
Cramer's week ahead: The market can go higher if Trump makes trade progress with Mexico
Cramer's lightning round: Tesla is just 'too darn hard' to recommend
Cramer Remix: I don't like the fossil fuel stocks
Cramer: Beware of froth — 'it may be too late to put new money to work'
Salesforce's Marc Benioff defends Tableau deal: They 'didn't want our cash'
Cramer's lightning round: Don't buy Roku here — it's due for pullback
Cramer Remix: The June Swoon talk is useless
The S&P 500 could fall to the 2,600 area, says strategist
Tableau CEO breaks down how the company will operate within Salesforce
The charts warn that stocks are due for a pullback, Jim Cramer says
Cramer's lightning round: Yeti is 'undervalued' and a 'long-time hold'
Cramer Remix: These stocks should be bought in this environment
Cramer: Hershey's turnaround came at the right time, and sent the stock surging 28%
Leading mortgage lender Quicken Loans sees record loan volumes
Cramer's lightning round: This stock is just too cheap. You should buy
Cramer Remix: Put new IPO Revolve on your shopping list
Cramer: Shares of Michael Kors' parent company are cheap — go with the bulls
Cramer's lightning round: Ball is in a duopoly. It's a 'total windfall'
Dexcom CEO teases a diabetes collaboration with Apple Watch
Cramer Remix: Thermo Fisher and Danaher are in a sweet spot to buy
Jim Cramer: Put Chewy on the shopping list and buy it on a pullback
Cramer's lightning round: Resmed had a hiccup. That was a chance to buy
Cramer Remix: I don't like this stock because of its exposure to US malls
The Shanghai index is making a 'major turnabout': Daryl Guppy
Jim Cramer shuffles his 'Humanization of Pets ETF' — 'It's paid off in 2019'
IFF CEO on the impacts of Brexit and China tariffs on business
Cramer Remix: What I'm expecting from the Fed on Wednesday
Tariffs hurt the economy, but China must be held accountable, Union Pacific CEO says
Starbucks and Dunkin' are 'defying the bears' — buy on a dip, Jim Cramer says
Cramer's lightning round: I think you should buy shares of Spotify
Cramer Remix: Libra is a 'genuinely positive thing' for Facebook
Cramer: The Fed pleased investors, but there's 'no escaping' the wrath of this White House
Cramer's lightning round: Signet Jewelers is still in the 'penalty box'
Cramer Remix: Adobe is the way to play the 'golden age of creativity'
Caterpillar CEO on China trade war fears: We're 'still meeting our forecast'
MongoDB CEO explains where the company has an edge over database giant Oracle
Cramer's lightning round: Lockheed Martin is the one you have to own
Cramer Remix: I think this pharma distributor is a house of pain
This could be 'the next shoe to drop' with stocks at all-time highs
2019-06-03
2019-06-03
2019-06-03
2019-06-03
2019-06-04
2019-06-04
2019-06-05
2019-06-04
2019-06-04
2019-06-05
2019-06-05
2019-06-05
2019-06-06
2019-06-06
2019-06-07
2019-06-07
2019-06-07
2019-06-07
2019-06-10
2019-06-10
2019-06-10
2019-06-10
2019-06-11
2019-06-11
2019-06-11
2019-06-11
2019-06-11
2019-06-12
2019-06-12
2019-06-12
2019-06-12
2019-06-13
2019-06-13
2019-06-13
2019-06-13
2019-06-14
2019-06-14
2019-06-14
2019-06-17
2019-06-17
2019-06-17
2019-06-17
2019-06-18
2019-06-18
2019-06-18
2019-06-19
2019-06-19
2019-06-19
2019-06-19
2019-06-20
2019-06-20
2019-06-20
2019-06-20
2019-06-21
@souravdey
0
50
100
150
200

Trending Stories Extracted



References

1. https://wordpress.com/activity/.
2. Identifying hidden trends in news stories using hierarchical clustering, https://towardsdatascience.com/identifying-hidden-trends-in-news-stories-using-hierarchical-clustering-b6297df795af
3. M. Tarik Altuncu, Sophia N. Yaliraki, and Mauricio Barahona. 2018. Content-driven, unsupervised clustering of news articles through multiscale graph partitioning. In Proceedings of 2018 KDD Data Science, Journalism and Media Text Classification in Python,