



r/WallStreetBets **Topic Modeling**

Manveer Sadhal
Nov 12, 2021



QUESTIONS



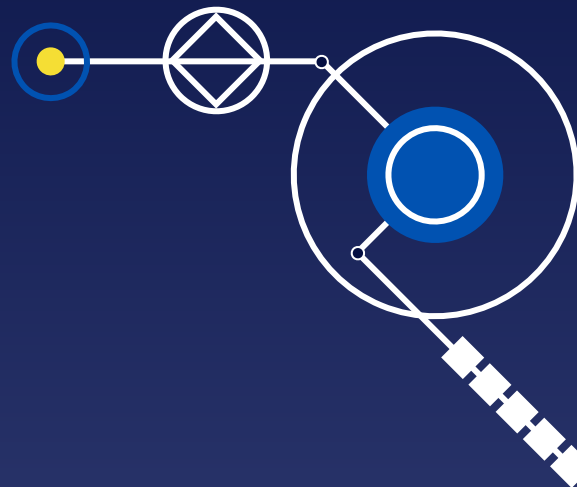
- What are the main topics of discussion on r/WallStreetBets?
- How do topics change over time?



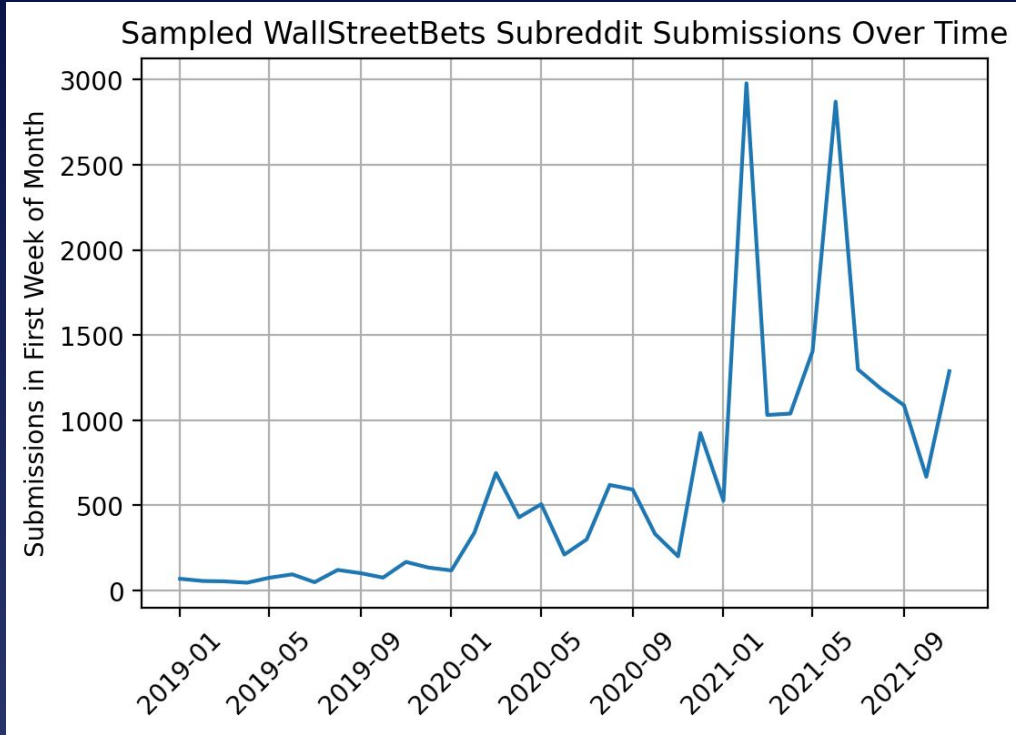
DATA AND SCOPE

- Text from r/WallStreetBets submissions
- “Deep dive” or “discussion” with length of at least 100 characters.
- ~25% of overall submissions in subreddit
- Jan 2019 to Nov 2021 - first week of month

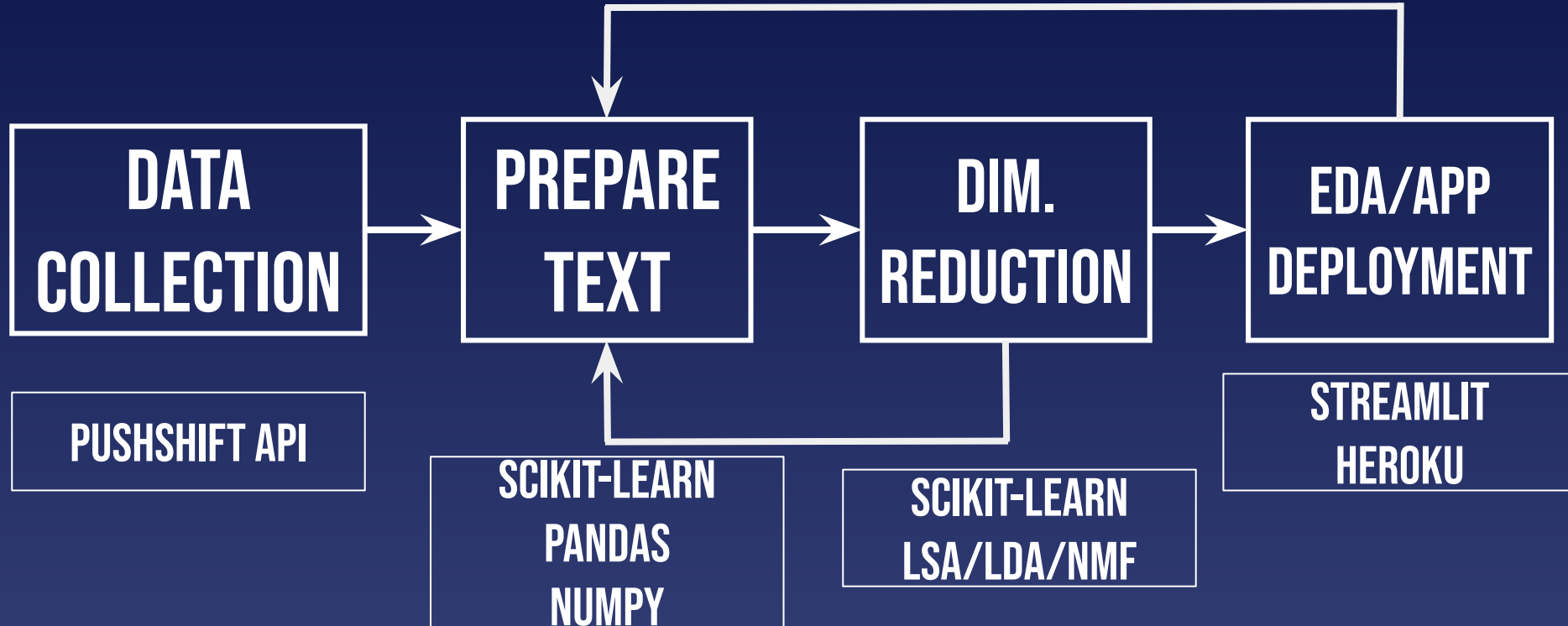
22,500
documents



POPULARITY OVER TIME



METHODOLOGY





RESULTS

NMF model with TF-IDF Vectorization produced the most coherent topics. Nov 2021 example below:

Topic	Keywords
Business results discussion	Growth, Potential, Quarter, Volume, News
Tesla and Elon Musk's net worth	Tesla, Elon, Car, Musk, Worth
GameStop short squeeze	GameStop, Nation, MOASS, State, Interested
Requests for investment advice	Help, Advice, Investing, Thank, Account
Tilray Cannabis, JPM restricting trades	Tilray, Cannabis, Citadel, Banking, JPM

WEB APP

<https://wsb-app-metis.herokuapp.com/>

WallStreetBets Topic Modeling

Choose Month, Model Type, and Model Parameters Below

Month to Model

November 2021

Type of Model

NMF

Maximum Features

903

Minimum Document Frequency

0.00

Number of Topics

7

Maximum Document Frequency

0.90

Words per Topic to Display

5

Model Results:

Topic 1: growth, potential, quarter, volume, news

Topic 2: tesla, elon, car, musk, worth

Topic 3: gamestop, nation, moass, state, interested

Topic 4: help, advice, investing, thank, account

Topic 5: tilray, cannabis, citadel, banking, jpm

Topic 6: vaccine, pfizer, fy2022, 120b, guidance

Topic 7: bbby, amc, float, borrow, increased

APP DEMO



CONCLUSIONS

- NMF topic modeling reveals insights about WSB discussions.
- Persistent themes over time:
 - Tesla and other EV-related companies
 - Computer tech (AMD, AAPL, NVDA, PLTR)
 - Short squeezes (e.g. GameStop, AMC)
 - COVID-19
 - Requests for investment advice
 - China

FUTURE WORK

- Collect full data set rather than sampling.
- Sentiment analysis relative to specific stocks.
- Correlate discussion with stock market moves.
- Incorporate outputs of topic modeling and sentiment analysis into stock price prediction algorithm.

THANKS!

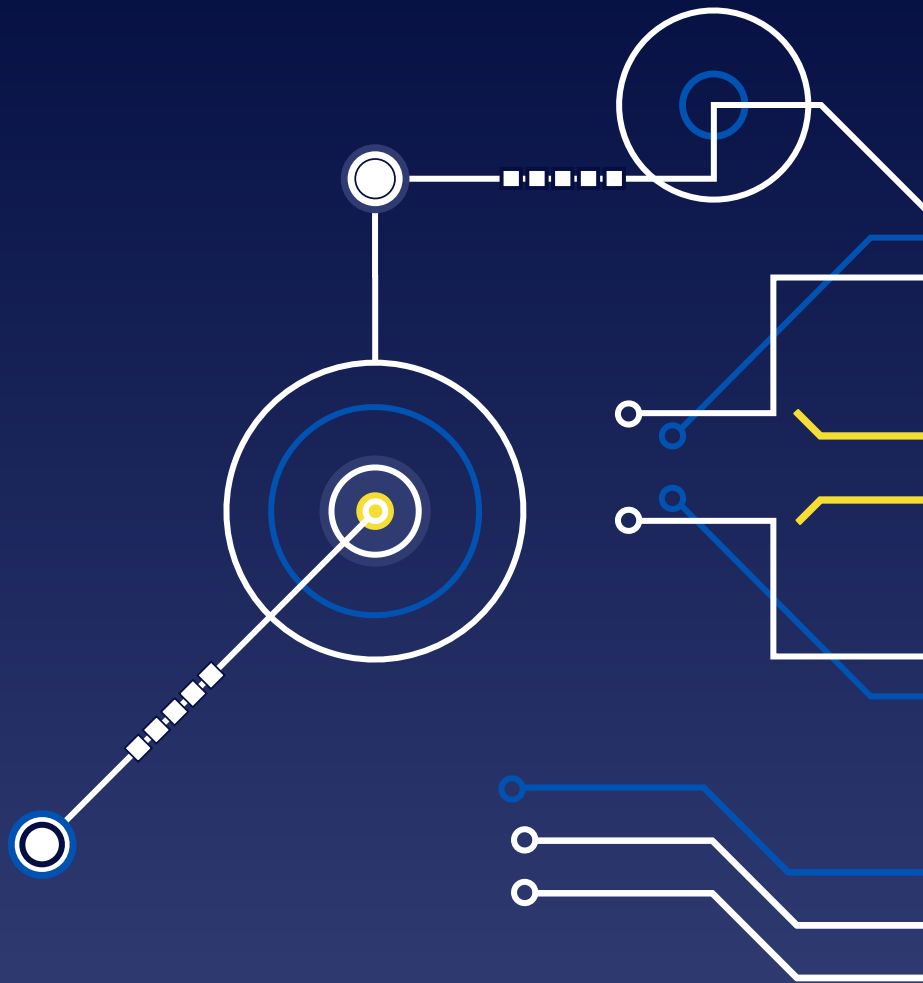
Do you have any questions?

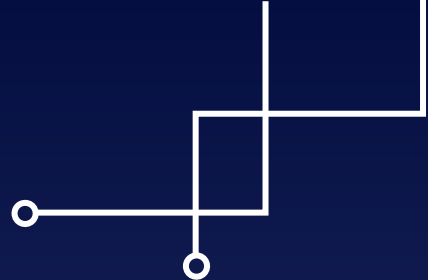
<https://github.com/manveersadhal>

<https://wsb-app-metis.herokuapp.com/>

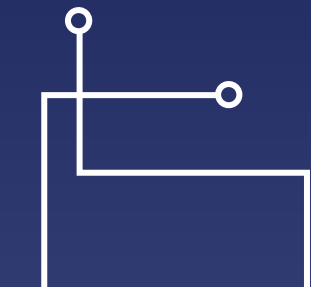
CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik

Please keep this slide for attribution





APPENDIX



☒ Word Clouds

☐ Topic Modeling

Word Cloud Comparison

[Click Here to Randomly Select Two Months to Compare](#)

First Month

April 2020



Second Month

November 2021



2019-01-01	[china, apple, markets, sales, guidance]	2020-07-01	[tesla, elon, musk, car, cars]
2019-02-01	[snap, app, public, bad, puts]	2020-08-01	[beat, covid, tomorrow, quarter, eps]
2019-03-01	[deal, beat, help, investing, china]	2020-09-01	[tesla, elon, battery, papa, inclusion]
2019-04-01	[puts, tomorrow, growth, daily, target]	2020-10-01	[tesla, news, hit, volume, worth]
2019-05-01	[tomorrow, try, morning, mid, playing]	2020-11-01	[biden, trump, wins, win, related]
2019-06-01	[puts, spy, risk, having, funds]	2020-12-01	[palantir, government, moon, puts, contracts]
2019-07-01	[rate, cut, margin, bad, higher]	2021-01-01	[growth, tech, energy, higher, positions]
2019-08-01	[puts, yolo, playing, worth, selling]	2021-02-01	[hedge, funds, volume, financial, silver]
2019-09-01	[puts, spy, volume, work, news]	2021-03-01	[rkt, rocket, late, distraction, dividend]
2019-10-01	[expected, tgt, yr, initiated, guidance]	2021-04-01	[volume, tesla, potential, value, news]
2019-11-01	[robinhood, free, leverage, financial, platform]	2021-05-01	[potential, news, term, growth, float]
2019-12-01	[robinhood, glitch, past, ceo, collateral]	2021-06-01	[amc, gamestop, hedgies, hedge, sold]
2020-01-01	[amd, ces, news, 7nm, contracts]	2021-07-01	[growth, volume, advice, float, news]
2020-02-01	[tesla, past, ago, spce, musk]	2021-08-01	[growth, help, amd, news, covid]
2020-03-01	[coronavirus, virus, cases, beer, china]	2021-09-01	[growth, term, financial, potential, quarter]
2020-04-01	[china, virus, covid, debt, coronavirus]	2021-10-01	[value, debt, term, news, volume]
2020-05-01	[chegg, term, covid, news, eps]	2021-11-01	[growth, potential, quarter, past, volume]
2020-06-01	[covid, sales, positions, tldr, recovery]		