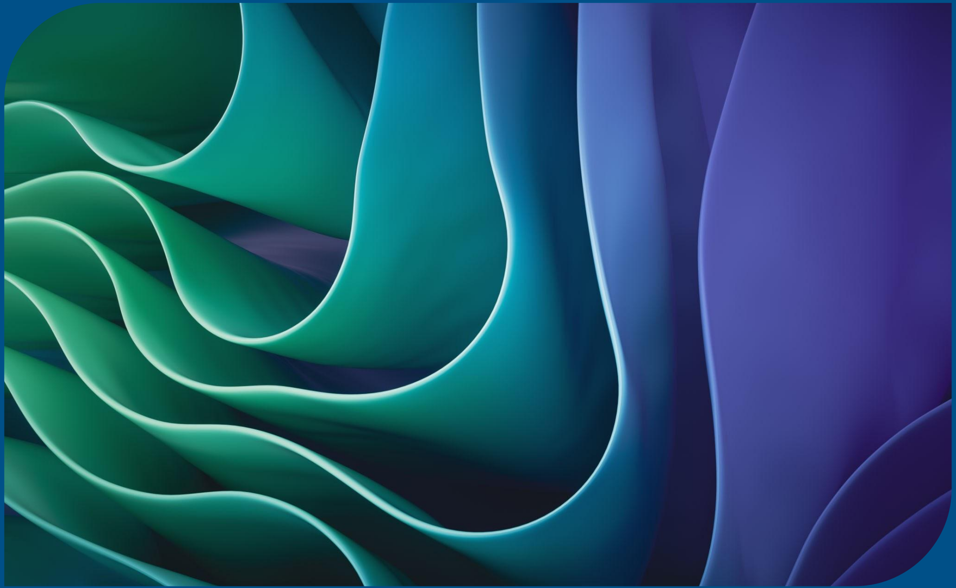
Tweet Sentiments Impact on Stocks



Group 17 Insight Innovators
Yuxuan Li (yl8095)
Siyuan Lu (sl10865)
Mason Lonoff (ml9542)

Research Questions

Main focus: *How do Tweet Sentiments Affect Stock Behavior?*

Subquestions:

1. Are stocks with high and low volatility affected differently by tweet sentiments?
2. How do tweets and their sentiment scores affect stocks across industries?
3. Which stocks are most sensitive to tweet sentiments?

Dataset

Tweet Sentiment's Impact on Stock Returns:
https://www.kaggle.com/datasets/thedevastator/tweet-sentiment-s-impact-on-stock-returns/data

Dataset consists of 14 columns and 1,395,450 rows

- **Tweet** - Text of the tweet. (String)
- **Stock** - Company's stock mentioned in the tweet. (String)
- **Date** - Date the tweet was posted. (Date)
- **LAST_PRICE** - Company's last price at the time of tweeting (Float)
- **1_DAY_RETURN** - Amount the stock returned or lost over the course of the next day after being tweeted about. (Float)
- **2_DAY_RETURN** - Amount the stock returned or lost over the course of the two days after being tweeted about. (Float)
- **3_DAY_RETURN** - Amount the stock returned or lost over the course of the three days after being tweeted about. (Float)
- **7_DAY_RETURN** - Amount the stock returned or lost over the course of the seven days after being tweeted about. (Float)
- **PX_VOLUME** - Volume traded at the time of tweeting. (Integer)
- **VOLATILITY_10D** - Volatility measure across 10 day window. (Float)
- **VOLATILITY_30D** - Volatility measure across 30 day window. (Float)
- **LSTM_POLARITY** - Labeled sentiment from LSTM. (Float)
- **TEXTBLOB_POLARITY** - Labeled sentiment from TextBlob. (Float)

1. Data Cleaning and EDA
2. Analyze sentiment impact on stocks with different volatility
3. How do tweets and their sentiment scores affect stocks across industries?
4. Determine the most sensitive stocks to tweet sentiment
5. Limitations

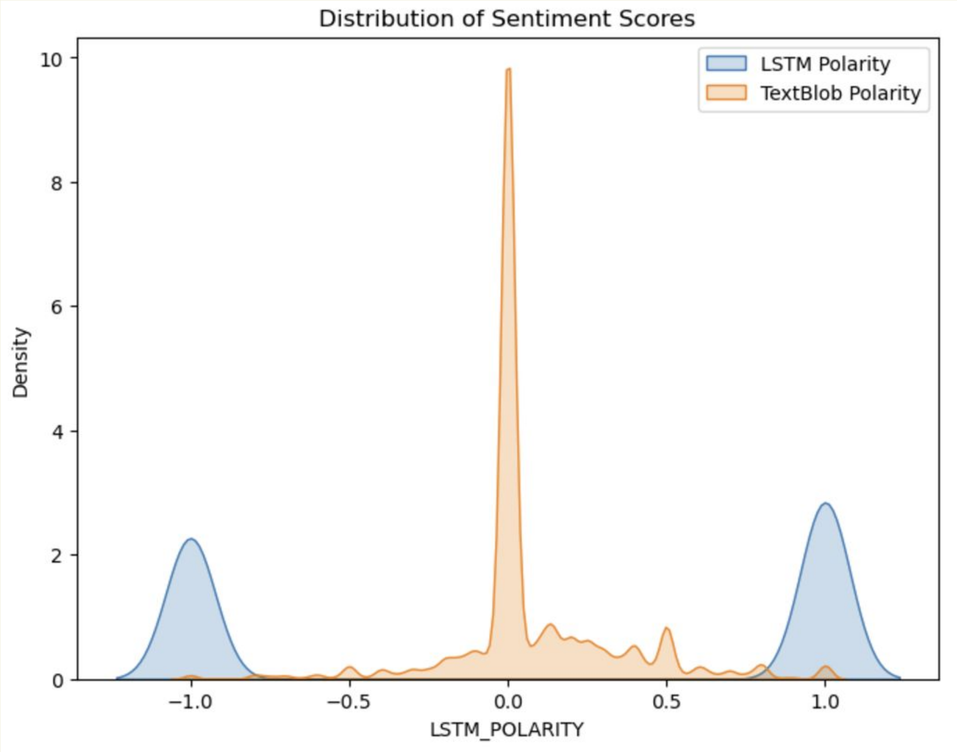
Data Cleaning and EDA

	Unnamed: 0	TWEET	STOCK	DATE	LAST_PRICE	1_DAY_RETURN	2_DAY_RETURN	3_DAY_RETURN	7_DAY_RETURN	PX_VOLUME	VOLATILITY_10D	VOLATILITY_30D	LSTM_POLARITY	TEXTBLOB_POLARITY
0	0	RT @robertoglezcano: @amazon #Patents Show Fl...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	Amazon	31/01/2017	823.48	0.008379	0.014924	0.014924	-0.001263	3.137196e+06	1.344700e+01	16.992	1.000	0.000000	NaN
2	1	@FAME95FM1 Jamaicans make money with @Payoneer...	PayPal	31/01/2017	39.780000	0.002011	0.012318	0.012318	5.480141e-02	9.100057e+06	18.769	16.099	-1.000000	0.0
3	2	@CBSi Jamaicans make money with @Payoneer @Pay...	PayPal	31/01/2017	39.780000	0.002011	0.012318	0.012318	5.480141e-02	9.100057e+06	18.769	16.099	1.000000	0.0
4	3	@Hitz92fm Jamaicans make money with @Payoneer ...	PayPal	31/01/2017	39.780000	0.002011	0.012318	0.012318	5.480141e-02	9.100057e+06	18.769	16.099	-1.000000	0.0
5	4	RT @loadsofvans: Retweet this post & follo...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

TWEET	STOCK	DATE	LAST_PRICE	1_DAY_RETURN	2_DAY_RETURN	3_DAY_RETURN	7_DAY_RETURN	PX_VOLUME	VOLATILITY_10D	VOLATILITY_30D	LSTM_POLARITY	TEXTBLOB_POLARITY
@FAME95FM1 Jamaicans make money with @Payoneer...	PayPal	31/01/2017	39.78	0.002011	0.012318	0.012318	0.054801	9100057.0	18.769	16.099	-1.0	0.0000
@CBSi Jamaicans make money with @Payoneer @Pay...	PayPal	31/01/2017	39.78	0.002011	0.012318	0.012318	0.054801	9100057.0	18.769	16.099	1.0	0.0000
@Hitz92fm Jamaicans make money with @Payoneer ...	PayPal	31/01/2017	39.78	0.002011	0.012318	0.012318	0.054801	9100057.0	18.769	16.099	-1.0	0.0000
RT @nikitakhara: Thank you, @Starbucks CEO for...	Starbucks	31/01/2017	55.22	0.012314	0.016298	0.016298	0.058312	14307985.0	23.916	17.298	1.0	0.2000
@gawker Jamaicans make money with @Payoneer @P...	PayPal	31/01/2017	39.78	0.002011	0.012318	0.012318	0.054801	9100057.0	18.769	16.099	-1.0	0.0000

1. Filter out rows where 'STOCK' is 'Next' but 'TWEET' does not contain '@nextofficial'.
2. Drop rows where 'Unnamed: 0' has missing values.
3. Reset the index of the DataFrame.
4. Drop the 'index' and 'Unnamed: 0' columns.
5. Drop any remaining rows with missing values in the entire dataset.

Data Cleaning and EDA



- **LSTM Polarity:** bimodal distribution

  LSTM model often classifies sentiments as strongly positive or strongly negative, with little to no neutral predictions.

- **TextBlob Polarity:** sharply peaked at 0, with a much smaller density in the positive and negative sentiment ranges

  TextBlob tends to assign neutral sentiment more often and is less confident in identifying strongly positive or negative sentiments.

- **Since TextBlob shows a large number of neutral sentiments, we focused only on the LSTM model in our following analysis.**

Data Cleaning and EDA

Correlation Matrix:

	LSTM_POLARITY	TEXTBLOB_POLARITY	1_DAY_RETURN
LSTM_POLARITY	1.000000	0.098295	0.004060
TEXTBLOB_POLARITY	0.098295	1.000000	-0.023269
1_DAY_RETURN	0.004060	-0.023269	1.000000
2_DAY_RETURN	-0.000302	-0.025690	0.785274
3_DAY_RETURN	-0.008917	-0.038014	0.646872
7_DAY_RETURN	-0.018221	-0.020226	0.418018

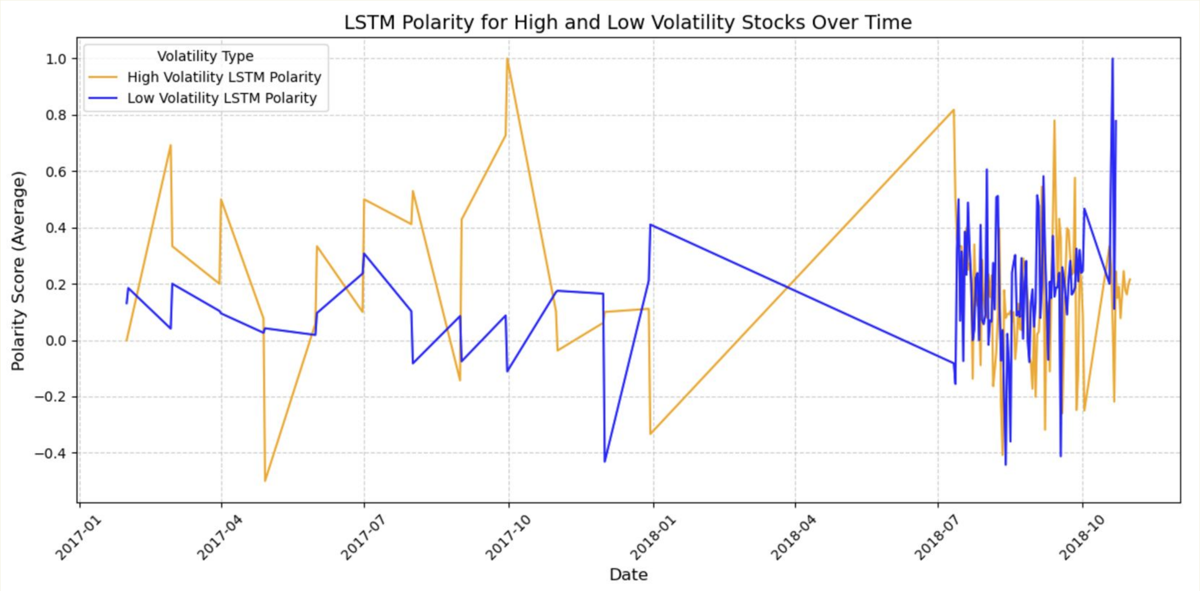
	2_DAY_RETURN	3_DAY_RETURN	7_DAY_RETURN
LSTM_POLARITY	-0.000302	-0.008917	-0.018221
TEXTBLOB_POLARITY	-0.025690	-0.038014	-0.020226
1_DAY_RETURN	0.785274	0.646872	0.418018
2_DAY_RETURN	1.000000	0.830636	0.571003
3_DAY_RETURN	0.830636	1.000000	0.679813
7_DAY_RETURN	0.571003	0.679813	1.000000

- LSTM and TextBlob results on sentiment scores are not strongly aligned.
- LSTM and TextBlob sentiment scores do not have a strong linear relationship with stock returns.

1. Data Cleaning and EDA
2. **Analyze sentiment impact on stocks with different volatility**
3. How do tweets and their sentiment scores affect stocks across industries?
4. Determine the most sensitive stocks to tweet sentiment
5. Limitations

Sentiment Impact on Stocks with Different Volatility

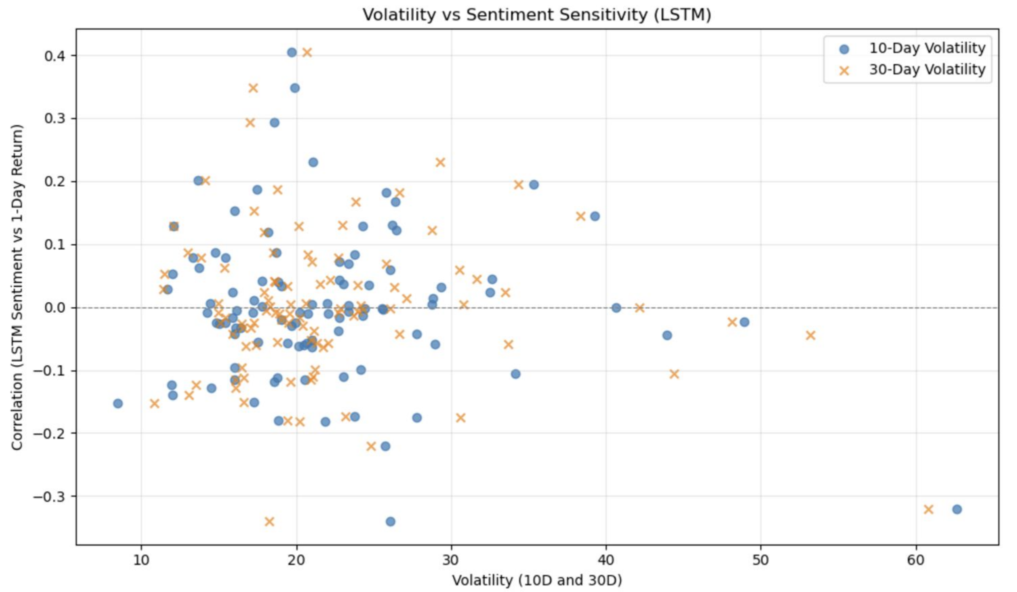
—- Sentiment trends over time



- The LSTM polarity scores for both high- and low-volatility stocks fluctuate significantly over time, we cannot observe any clear patterns suggesting that high or low volatility is related to LSTM sentiment polarity through this analysis.

Sentiment Impact on Stocks with Different Volatility

—- Correlation analysis of LSTM sentiment

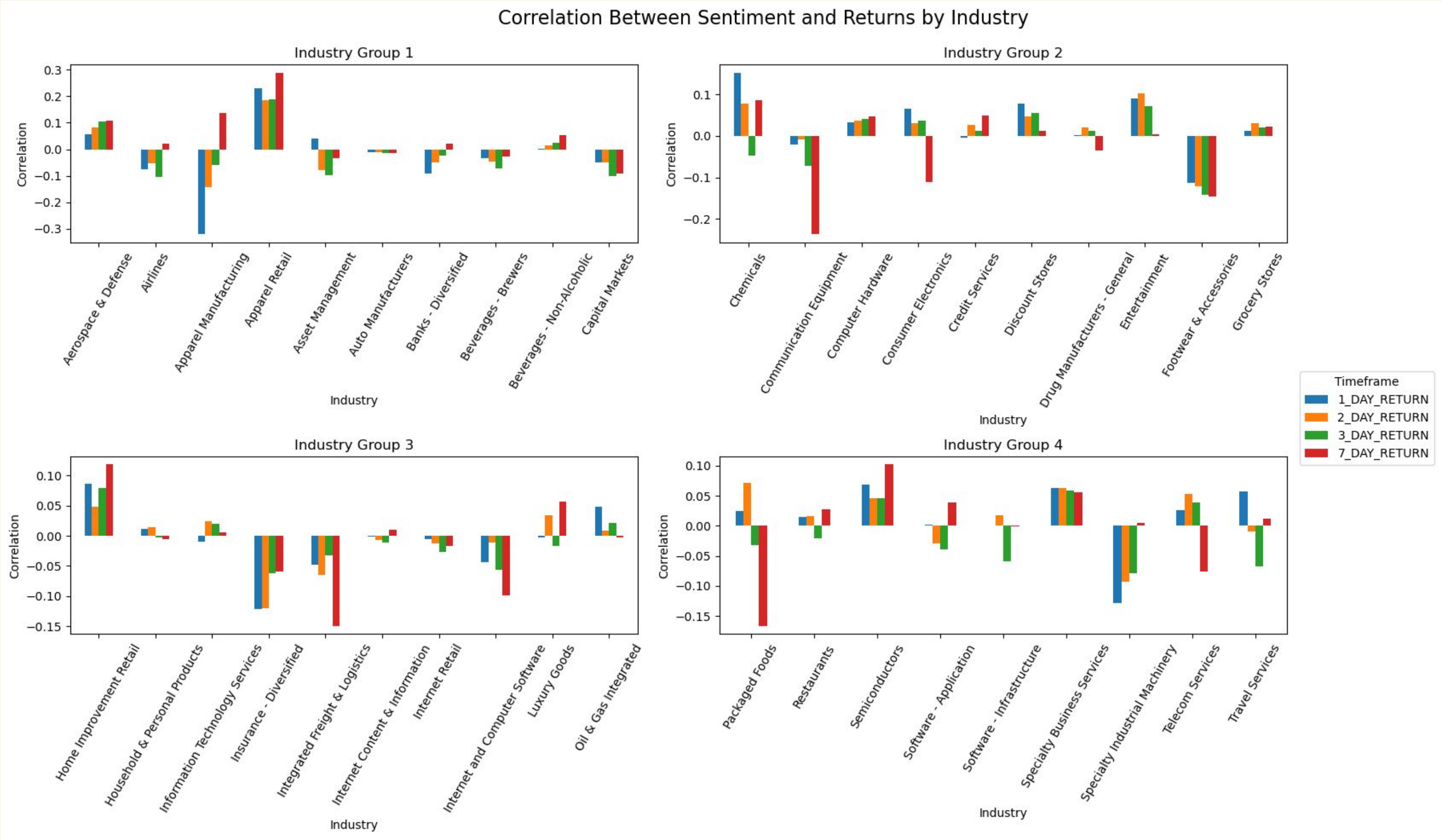


	STOCK	VOLATILITY_10D	VOLATILITY_30D	LSTM_CORRELATION
45	H&M	62.635739	60.825516	-0.320622
44	Groupon	48.903412	48.134014	-0.022620
95	Yahoo	43.936898	53.216219	-0.043839
36	Facebook	40.643031	42.153120	-0.000698
64	Netflix	39.238724	38.375806	0.144081

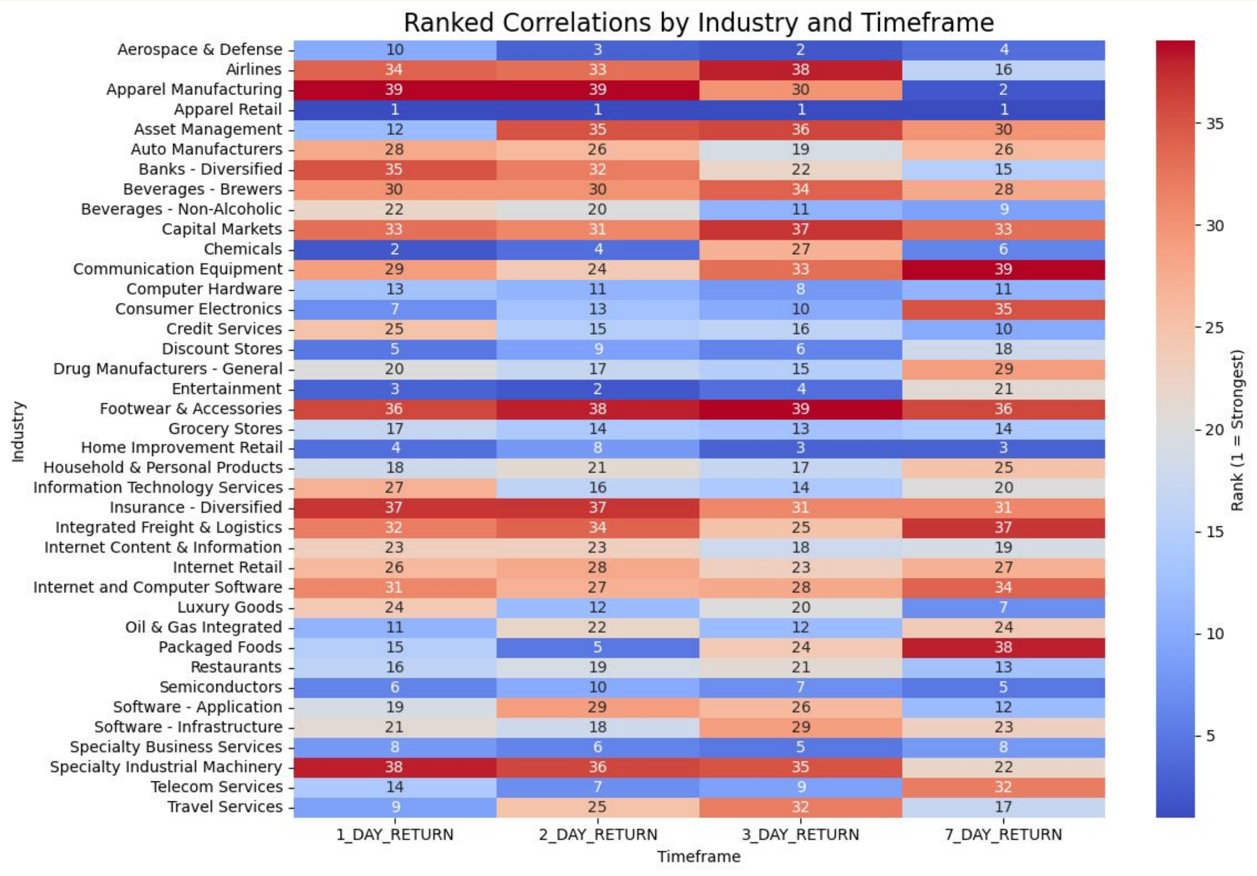
- No strong linear relationship between volatility (both 10-day and 30-day measures) and sentiment sensitivity (correlation between LSTM polarity and 1-day returns).
- Stocks with low volatility clusters around low sentiment sensitivity.
- High volatility doesn't consistently translate to strong sensitivity to tweet sentiments.

1. Data Cleaning and EDA
2. Analyze sentiment impact on stocks with different volatility
3. **How do tweets and their sentiment scores affect stocks across industries?**
4. Determine the most sensitive stocks to tweet sentiment
5. Limitations

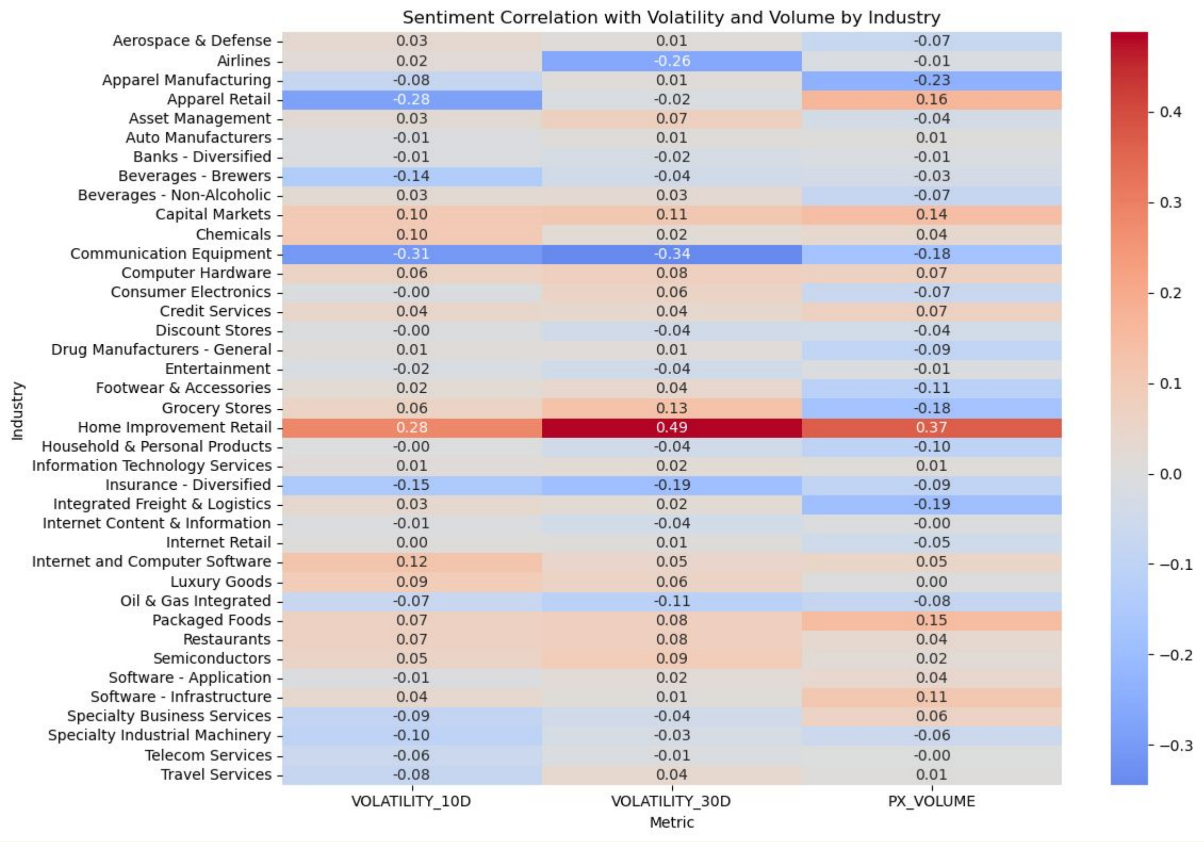
Correlations of Different Return Timeframes with Sentiment Scores



How Does Each Industry Compare to Eachother?

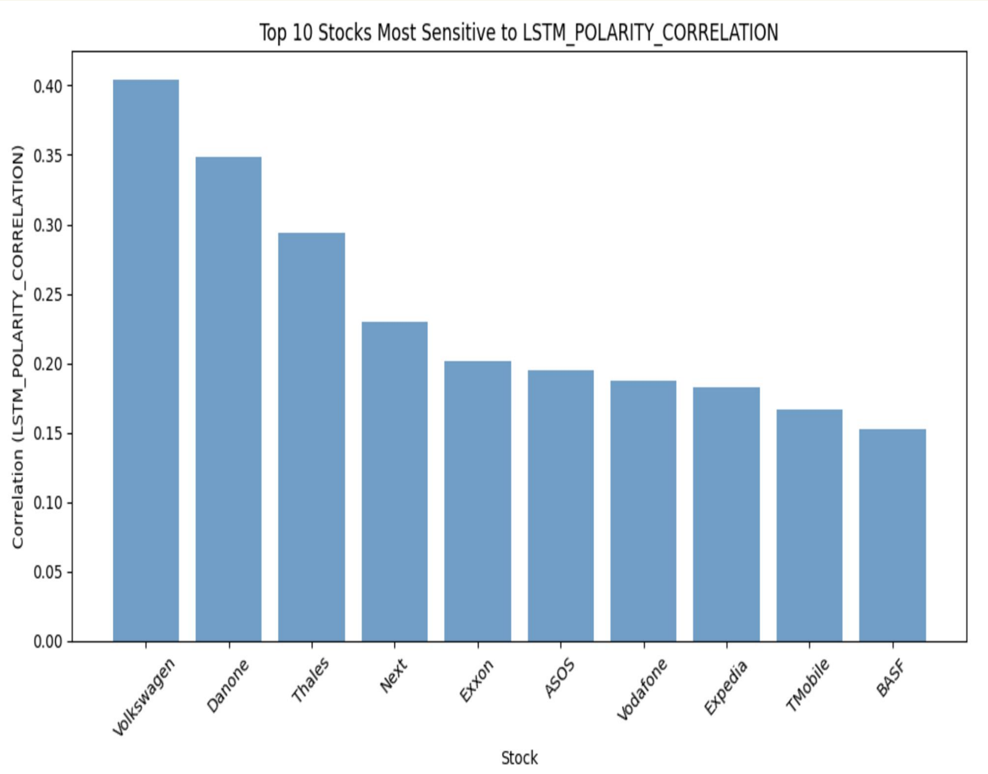


What’s the Relationship Between Sentiment Scores with Volatility and Volume?



1. Data Cleaning and EDA
2. Analyze sentiment impact on stocks with different volatility
3. How do tweets and their sentiment scores affect stocks across industries?
4. **Determine the most sensitive stocks to tweet sentiment**
5. Limitations

Determine the Most Sensitive Stock to Tweet Sentiment

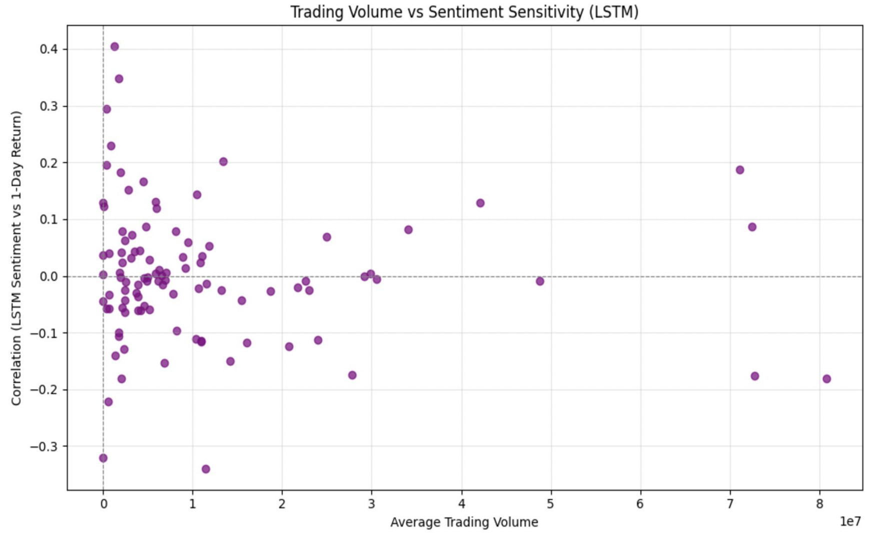


- Industry Matters
- Investor Behavior
- Potential Practical Implications

Determine the Most Sensitive Stock to Tweet Sentiment

—- Volume-Traded Perspective

Assumption: Stocks with higher trading volumes might exhibit greater sensitivity to tweet sentiments because they attract more traders who react to public information.



	STOCK	AVG_VOLUME	LSTM_CORRELATION
77	Santander	8.076309e+07	−0.180441
40	General Electric	7.277066e+07	−0.175711
14	Bank of America	7.247274e+07	0.087381
91	Vodafone	7.112379e+07	0.187255
38	Ford	4.878010e+07	−0.009418

- Weak Overall Relationship
- High-Volume Stocks with Positive Correlation
- Contrarian Behavior in High-Volume Stocks
- High-Volume Stocks with Weak Correlation

1. Data Cleaning and EDA
2. Analyze sentiment impact on stocks with different volatility
3. How do tweets and their sentiment scores affect stocks across industries?
4. Determine the most sensitive stocks to tweet sentiment
5. **Limitations**

Limitations

- **Inconsistent Sentiment Scores for the Same Tweet**
  Different sentiment analysis models can produce varying scores for the same tweet, leading to inconsistencies and conflicting results. Some identical tweets also resulted in differing scores
- **Timeframe of Tweets and Specific Market Conditions**
  Tweets are collected over different time spans for each stock, and market conditions (e.g. volatility) vary across these timeframes.
- **Subsidiary Relationships Between Companies**
  Some companies in the dataset are subsidiaries of large parent companies. Tweets mentioning the parent or subsidiary may indirectly impact the other, causing correlation overlap.
- **Quarterly Differences in Data Volume**
  The number of tweets available for analysis may vary significantly across different quarters due to seasonal trends or reporting periods.

Questions?