Strength in Numbers

Reddit, Returns, and Rhetoric: Retail Investor Impact on Technology Stocks in a Politically Charged Market

JOHAN HANNEGÅRD

AXEL KARLSSON

Bachelor Thesis

Stockholm School of Economics

2025

Abstract
This thesis investigates how retail investor sentiment on Reddit influences the performance of four big tech stocks (Tesla, Nvidia, Microsoft, and Apple) during a politically charged period. The analysis uses the Valence Aware and sEntiment Reasoner (VADER) for sentiment analysis, paired with daily stock data to capture the relationship between retail sentiment and stock returns, as well as trading activity. The study finds that sentiment has a generally weak and statistically insignificant correlation with daily stock returns of the chosen companies. However, the findings reveal stronger and more consistent relationships between sentiment and trading volume, particularly for Tesla, suggesting that sentiment may act as a behavioral trigger without consistently affecting prices. The study contributes to understanding the nuanced role of social sentiment in financial markets, particularly during periods of heightened political uncertainty.

Keywords: Valence Aware Dictionary sEntiment Reasoner, Big Tech, Pearson Correlation, Efficient Market Hypothesis

Authors:
Johan Hannegård (25726)
Axel Karlsson (25720)

Tutor:
Marieke, Bos, Deputy Director, Swedish House of Finance

Acknowledgements:
Firstly, we would like to extend our deepest gratitude to Marieke Bos, our supervisor, for her vital guidance and constructive feedback throughout the research process. We would also like to thank our peers in the supervision group for their though-provoking discussions and continuous support. Finally, we express our sincere appreciation to Axel Tardell for his invalubale insights specifically with regards to his expertise in mathematical modeling.

Bachelor Thesis
Bachelor Program in Business and Econoomics
Stockholm School of Economics

1. Introduction

1.1 Motivation and Purpose

Markets are not always efficient. While the Efficient Market Hypothesis (EMH) in traditional finance theory assumes that all available information is fully reflected in asset prices, recent events show that retail investors, especially those on social media, can move markets in ways that defy these fundamental beliefs. A good example is the GameStop short squeeze in 2021, where Reddit users pushed prices beyond what many believed was justified by earnings. This case exemplifies that online sentiment can play a major role in market behavior, underscoring its influence during one of the most prominent and widely discussed market phenomena in recent years (Salvucci, 2023).

According to a report by the CFA Institute, 60% of U.S. investors under the age of 35 use social media as a source of investment information. Similarly, investors under 34 are the most likely group to trust social media platforms, such as Reddit, TikTok, and YouTube, for financial insights (Preece, 2024). These trends underscore social media's growing role as both a sentiment hub and a source of trading activity, and supports the growing relevance of Reddit as an investment platform. The subreddit r/WallStreetBets, for example, grew from roughly 100,000 subscribers in 2016 to over 13 million by the end of 2022, reflecting Reddit's rise as a central node for market narratives (Nugent, 2023). Reactions from institutional investors are also evident. According to a 2023 survey by Brunswick and Institutional Investor, 58% of buy-side managers reported making an investment decision based on information from Reddit, and 49% rated Reddit as a source of 'high importance' for stock valuation. However, only about 23% of institutions report actively harvesting data from Reddit (via NLP/AI) for analysis, indicating that the majority that use Reddit as a source of investment analysis are retail investors (McElhaney, 2023). This group of investors has also significantly increased in size. Industry reports estimate that around 30 million new U.S. retail investors opened brokerage accounts in 2020, and by 2021 retail traders accounted for 25% of total U.S. equity trading volume (nearly double their share a decade earlier). Even into 2023, retail cash inflows remained at record levels, e.g. over $1.5 billion into U.S. equities in one week of February 2023 (Einhorn et al., 2023).

The rapid expansion of both young investors, retail investing, and Reddit usage creates an ideal environment for studying the intersection of sentiment and market behavior. We aim to explore whether political news during intense political periods contributes to shifts in investor sentiment that may influence market behavior beyond what fundamental analysis would predict as outlined by the EMH (Efficient Markets Hypothesis). By tracking how Reddit users react during a politically charged time-frame; defined as a period characterized by heightened political discourse, significant policy announcements, or contentious political events that capture public attention, we aim to capture how emotional or belief-driven responses can influence actual trading patterns and stock prices. This focus is further driven by the growing importance of social media as a bridge between political discourse and investment decisions. Politicians, most notably Donald Trump, frequently use platforms like

X to communicate directly with the public, often employing provocative or emotionally charged language. Such communication not only amplifies polarization but also fuels instantaneous, emotionally driven reactions among users (Sugihartono, 2024). For retail investors who gather on Reddit, this dynamic creates an environment where breaking political developments can quickly influence collective sentiment and, in turn, market behavior. In light of this, we formulate the following research question: *How does retail investor sentiment on Reddit, during a period of heightened political tension, influence technology stocks that are prominent in both media and regulatory attention?* Our hypothesis is: Retail investor sentiment on Reddit has a significant impact on the daily returns and trading volumes of selected big tech stocks (Tesla, Nvidia, Microsoft, and Apple).

To explore this, we analyze four technology stocks: Tesla, Nvidia, Microsoft, and Apple. These companies were selected because they are not only among the most frequently discussed on Reddit, but also because they represent a spectrum of market behavior within the technology sector (Reddit, 2024). Tesla and Nvidia are known for their high volatility and sensitivity to public sentiment, often responding sharply to news related to innovation, leadership, and regulation. In contrast, Microsoft and Apple are more stable and mature firms with diversified business models, but they are still exposed to politically sensitive issues such as data privacy, antitrust, and global trade. By including both more volatile and relatively stable technology firms, we aim to capture a broader understanding of how retail sentiment on Reddit influences stock performance under the set time period. This focus also allows us to test our hypothesis within a sector known for its rapid responses to public discourse.

1.2 Irrational Market

Understanding the role of sentiment is important because it shows that markets do not always move based on logic or fundamentals, which creates both opportunities and consequences. If prices are driven by emotion, this goes against key financial theories like the EMH, and stocks would be mispriced compared to their fundamentals. This paves the way for new research areas, such as finding ways to measure and model emotions in financial markets (Moody's, 2024). This has already been proven in the GameStop episode, which has led to the finance industry changing how it looks at the market. In the past, investors primarily focused on traditional financial metrics, such as company reports, earnings data, and macroeconomic indicators, to inform their decisions. However, contemporary investment strategies increasingly incorporate the analysis of online discourse, leveraging advanced tools such as natural language processing (NLP) to systematically assess and interpret public sentiment across digital platforms (Mercanti, 2024). This shift affects not just professionals in finance, but also people around the world whose savings and investments depend on the stock market.

1.3 Contribution

This thesis contributes to the existing literature by extending the work of Cookson and Niessner (2020), who established that belief-driven disagreement among retail investors persists even with new information due to confirmation bias and group thinking. Our study

builds on this framework by focusing on how investor disagreement evolves during a politically charged period, specifically examining whether sentiment polarization on Reddit intensifies its effect on the chosen stocks. Methodologically, we differ from Cookson and Niessner by using Reddit, a platform characterized by free-form, text-based discussions, instead of StockTwits, where sentiment is user-labeled (bullish/bearish). By combining this with Valence Aware Dictionary and sEntiment Reasoner (VADER), we are able to capture a broader and more nuanced spectrum of sentiment expressions. This approach reflects the emotional and behavioral complexity of real-world discussions, which cannot be fully represented through predefined sentiment categories. Our findings offer insights into the behavior of retail investors by demonstrating that sentiment can become polarized during periods of political uncertainty, even if the sentiment itself is not explicitly political. This suggests that disagreement among investors can be influenced by heightened public discourse, affecting their emotional responses and investment decisions regardless of the sentiment's thematic focus. This thesis adds to the understanding of how sentiment can influence trading activity, and price fluctuations, in particular technology stocks that attract significant retail attention, during a period of heightened political uncertainty.

2. Literature and Theoretical Framework

This section provides a comprehensive overview of the existing literature and theoretical foundations that inform our study. It begins by exploring the interplay between social media sentiment, retail investor behavior, and political uncertainty. Subsequent sections introduce VADER sentiment analysis, the EMH, investor sentiment, and the role of attention as a potential distortion to EMH. Finally, the section outlines the methodological considerations of our study design.

2.1 Literature Review

Several strands of research are relevant to our topic, but they are rarely connected. On one side, there is growing literature on social media sentiment and retail investor behavior. On the other, political uncertainty has long been studied in relation to market volatility. However, few studies explore how these two areas interact.

Cookson and Niessner (2020) provide the core theoretical base for our study. They show that disagreement among retail investors does not disappear when new information arrives. Instead, investors often stick to their prior beliefs and follow those who share similar views. Their study highlights how confirmation bias and social dynamics on platforms like StockTwits lead to persistent belief-driven disagreement. This idea is central to our hypothesis and our study builds on this by testing whether these dynamics become even stronger during politically charged periods. If the polarization of sentiment increases during a politically eventful period, it would suggest that investor behavior online is not only shaped by financial views, but also by political or personal beliefs. This would strengthen the Cookson and Niessner model by showing that it applies even outside normal financial settings. On the other hand, if sentiment stays stable or becomes less divided, it might mean

that retail investors are more focused on fundamentals than expected, or that political news does not have a big effect on trading behavior on Reddit. In either case, our findings have the potential to alter how the financial industry uses sentiment data to understand stock movements and market risk. It could also help institutional investors and platform moderators better understand what drives online volatility, and could lead to new research on how politics influences investor behavior.

Our work is also motivated by the observation that social media plays a growing role in both politics and investing. As mentioned, today's political headlines are often delivered through social media platforms like Instagram or X, with Trump using X as one of his primary communication tools. These platforms often generate emotionally charged reactions. This increases polarization, which may translate into stronger and more divided investor reactions. What makes this especially relevant is that many retail investors now gather on platforms like Reddit, where they read and react to news as it breaks (Sugihartono, 2024). This constant loop-flow of reactions means that news is not just read and forgotten. Instead, it is quickly shared, discussed, and often twisted to fit people's opinions or emotions. These reactions might not be based on financial logic, but they can still lead to real buying and selling in the market (Vantage, 2025).

In addition to Cookson and Niessner, other relevant studies include Semenova and Winkler (2021) and Lyócsa et al. (2021), who examine how retail sentiment on Reddit can create large, fast price swings. Both studies focus on the GameStop episode, where coordinated action by retail traders led to short-term speculative surges. Their work shows that collective sentiment can override fundamentals, at least temporarily, and that Reddit is a powerful amplifier of investor behavior. However, their focus is on isolated events in stable political periods, not on how this mechanism interacts with political news. Conversely, Pastor and Veronesi (2012) and Baker et al. (2016) show that elections and policy risk affect stock market volatility and investor uncertainty. They highlight that political cycles shape financial markets, especially through channels like regulation, taxation, or trade policy. However, these studies primarily focus on aggregate market reactions or institutional investors, rather than retail sentiment expressed online on platforms like Reddit. This leaves a gap in the literature: while we know that political risk matters, and that retail sentiment matters, we do not yet understand how the two interact. Our study addresses this gap by bringing these two lines of research together.

2.2 Attention as a Distortion to the EMH

Investor attention has long been discussed in behavioral finance as a driver of market movement, especially among retail traders. Da, Engelberg, and Gao (2011) show that increased online attention, measured via Google search volume, can predict stock returns and trading volume, particularly when investor attention is concentrated on a few high-profile stocks. In our case, Reddit acts as the central hub for attention. When political news is affecting popular stocks, and it leads to discussions online, we expect a spike in investor engagement, leading to changes in both volume and price volatility. This helps explain why certain companies, such as Tesla and Nvidia, may be more reactive. Their prominence on

social media platforms means they serve as “attention magnets,” particularly when linked to broader political themes like tariffs or regulation.

2.3 Trading Volume as Sign of Behavioural Finance

Research in behavioral finance suggests that social-media driven investor sentiment tends to stimulate trading activity without equal price effects. In other words, sentiment from platforms like Reddit can push more people to trade (extensive margin) even if it does not significantly move prices (intensive margin). Antweiler and Frank (2004) analyzed early internet stock message boards and found that online discussion postings predict increased trading volume and volatility, but have only a negligible impact on stock returns. This indicates that chatter on forums gets investors to trade more, yet any effect on prices is statistically small. In addition, Munster et al. (2024), studying Reddit posts vs. news articles, observed that both types of media trigger a short-lived surge in retail stock holdings (i.e. more Robinhood traders buying a stock), with social-media posts having a larger effect. Notably, investors flock to stocks mentioned on Reddit regardless of whether the post sentiment is positive or negative, highlighting an attention-driven trading spike rather than a fundamentally driven price change. These findings suggest that social sentiment is more likely to trigger trading activity than to reflect a stock’s true value. In summary, online buzz tends to increase market participation without materially altering asset prices in the long-run. This aligns with the idea that Reddit-hype can trigger trades even if it does not fundamentally reprice the stock.

2.4 Valence Aware Dictionary and sEntiment Reasoner (VADER)

VADER is a widely used lexicon and rule based sentiment analysis model originally introduced by Hutto and Gilbert (2014) to address the nuances of social media text. It was built by empirically deriving a sentiment lexicon attuned to informal language (e.g. slang, abbreviations, emojis) and defining five heuristic rules (for punctuation, capitalization, degree modifiers, negation, etc.) that account for how sentiment intensity is expressed in context. This approach produces interpretable sentiment metrics, each text is scored for positive, neutral, and negative sentiment, as well as an overall compound sentiment score normalized to range from -1 to +1 (Reichenbach et al., 2023). Importantly, VADER is simple to use but still provides accurate results. On benchmark evaluations with Twitter data, it performed on par with or better than more complex machine learning models, even outperforming individual human rates in distinguishing positive vs. negative sentiment (F1=0.96 vs 0.84) and demonstrating strong generalizability across domains. According to Pano et al (2020) a key advantage of VADER is its interpretability and domain-independence, since its lexicon and rules are human-understandable and not learned from domain-specific data, it can be applied to new datasets without retraining. Furthermore, unlike traditional bag-of-words models, which ignore context, VADER applies a set of rule-based grammatical and syntactic heuristics to account for the affective subtleties of informal online language. These rules correct and amplify the base valence score of each token depending on contextual

factors. For example, the presence of all-uppercase words ("AWFUL") and repeated punctuation ("!!!") increases sentiment intensity; degree modifiers such as "very" or "slightly" proportionally scale the base valance; and negation structures like "not good" invert or dampen the intended sentiment. Additionally, VADER incorporates a contrastive conjunction heuristic that places greater weight on the clause following a term like "but," thereby reflecting how human readers typically assign priority in sentiment interpretation (e.g., “The stock fell, but it recovered quickly” skews positive overall). However, being lexicon-based, VADER also has limitations. It may miss subtle context or sarcasm in text, and its general lexicon can fall short on domain specific jargon or slang (Malhotra et al., 2025). Overall, VADER offers a robust and transparent baseline for sentiment analysis, especially suited for social media data, making it a relevant tool for this thesis’s examination of Reddit and financial text sentiment, while acknowledging the need for slight domain adaptation to capture niche vocabulary.

2.4.1 VADER Mathematical Framework

VADER is employed to conduct sentiment analysis on Reddit data, enabling the measurement of user sentiment over time. VADER’s sentiment score calculation is grounded in a systematic approach that can be formally represented using the following equations (Hutto et al., 2014).
The raw sentiment score calculation is:

$$S = \sum_{i=1}^{n} (w_i \, x \, m_i \, x \, e_i)$$

The raw sentiment score is further normalized to produce a compound sentiment score, ensuring that the value falls between -1 and 1. The compound score normalization is represented by the following equation:

$$Compound\ Score = \frac{S}{\sqrt{S^2 + \alpha}}$$

Table 1: Variables in VADER Sentiment Scoring

Variable	Description	Determination Method
w_i	Sentiment intensity of a word	Pre-assigned in VADER lexicon
m_i	Modifier for intensity	Determined by presence of intensifiers or negations (e.g. “very”)
e_i	Emphasis multiplier	Adjusted by punctuation (“!”) or all capitalization
n	Number of sentiment-bearing words	Count of sentiment words in text
α	Normalization constant	Set to 15 (empirically determined)

2.5 Efficient Market Hypothesis (EMH)

The EMH, as formalized by Fama (1970), posits that asset prices at any point in time fully reflect all available information. Under this framework, financial markets are considered informationally efficient, meaning that new public information is rapidly and accurately incorporated into stock prices through the actions of rational, profit-maximizing investors. The EMH is commonly divided into three forms, weak, semi-strong, and strong, each representing a different level of information integration. The weak form suggests that all past trading information, such as historical prices and volumes, is already reflected in current prices. The semi-strong form, which is most relevant for studies involving news and sentiment, holds that all publicly available information, including earnings reports, macroeconomic data, and political developments, is instantaneously priced in. The strong form extends this logic further by asserting that even private or insider information is reflected in market prices.

This theoretical relationship is often expressed mathematically to illustrate how asset prices incorporate information. A common representation of the EMH is:

$$P_t = E\left[P_{t+1} | I_t\right]$$

Where P_t is the current price of the asset, and $E\left[P_{t+1} | I_t\right]$ i: the expected future price based on all information available at time t (I_t). This expression reflects the idea that today's price reflects all known information, and that future price changes are driven only by new, unexpected information. The expectation is conditional, meaning it explicitly depends on the information set I_t, and this is known as conditional probability. Another way to express this in terms of returns is:

$$R_{t+1} = E\left[R_{t+1} | I_t\right] + \epsilon_{t+1}$$

Here R_{t+1} is the return, and ϵ_{t+1} is an unpredictable error term capturing the arrival of new information. The error term has an expected value of zero, and is thus uncorrelated with the information set I_t, ensuring it captures only unexpected events. Under this view, price movements are essentially random:

$$P_{t+1} = P_t + \epsilon_{t+1}$$

Together, these equations formalize the idea that prices follow a random walk and are only affected by new, non-forecastable information, an implication central to the Efficient Market Hypothesis. This is a driftless random walk, meaning there is no constant growth term.

In the semi-strong form, the implication is that no investor can consistently earn abnormal returns by trading on public news or events, since the market has already adjusted for that information the moment it became available. This rests on assumptions of rational expectations, widespread access to information, and the existence of arbitrageurs who exploit and thereby eliminate mispricings. Within this framework, stock prices follow a random walk, and movements are primarily driven by the arrival of new, unexpected information that alters a firm's perceived value. As such, emotional reactions, investor sentiment, or public discourse, unless they convey new and material information, should have no persistent effect on asset prices. If the EMH holds in practice, political news would only influence stock prices to the extent that it directly alters expectations about firms' future cash flows or risk profiles, and any price adjustment would occur swiftly and rationally.

2.6 Comparative Analysis Theory

The Pearson correlation is a statistical measure that quantifies the linear relationship between two linear variables. The Pearson correlation coefficient ranges from -1 to 1, where values closer to plus, minus 1 indicate stronger linear relationships, and a value near zero suggests little to no linear association. The equation for Pearson correlation is as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

The accompanying p-value is a conditional probability that quantifies the likelihood of observing the given correlation (or a more extreme one) under the assumption that there is no true relationship between sentiment and returns, i.e., the null hypothesis holds. A 95% confidence level was applied, meaning that p-values below 0.05 are interpreted as statistically significant and suggest evidence of a meaningful linear relationship. p-value follows the equation:

$$p - value = P(TS\ ts \mid H_0\ is\ true) = cdf(ts)$$

3. Methodology

3.1 Research Design

This study investigates the relationship between political news, retail investor sentiment on Reddit, and short-term stock movements in four major tech companies. Given the dynamic and time-sensitive nature of online investor discussions, a quantitative design using sentiment analysis and financial market data was deemed appropriate. The research builds on prior work by Cookson and Niessner (2020), who highlight the persistence of investor disagreement in online communities. Their findings inform our decision to focus on retail sentiment and its potential divergence during politically charged periods. By extending this perspective into a political context, we aim to examine whether such disagreement intensifies

in response to major political events. To explore this, we selected a case window immediately following the 2024 U.S. presidential election, a period characterized by heightened political uncertainty. The study follows a structured observational design. Reddit posts were collected from r/stocks, r/investing, and WallStreetBets using the PRAW API. Sentiment was quantified using VADER, and combined with daily stock return and trading volume data sourced from Yahoo Finance. This approach allows us to align retail sentiment with financial performance on a daily basis.

To identify the effect of political news on retail sentiment and market behavior, this study relies on aggregated Reddit posts and stock-level data. While user-level data with timestamp trading actions directly linked to sentiment expressions, encompassing labeled sentiment (bullish/bearish), granular trading behavior, and demographic characteristics of retail traders would offer even greater precision, our chosen approach still provides valuable insights. Specifically, it allows us to capture broad patterns and correlations over time, reflecting the collective sentiment and its impact on market behavior. A controlled event-study framework with randomized political stimuli could further isolate causality. Since this level of granularity was not accessible, we instead rely on aggregated Reddit posts and stock-level data. This deviation limits our ability to make causal inferences and instead confines us to identifying patterns and correlations over time. In addition to this, we would also have liked to look at a longer time-frame, as this would allow us to examine whether the observed sentiment-market dynamics persist beyond the immediate post-election period and to distinguish between short-lived emotional reactions and more sustained behavioral trends. Compared to Cookson and Niessner (2020), who used StockTwits data containing user-tagged sentiment (e.g., “bullish” or “bearish”) and a broader cross-section of stocks, our dataset differs in two key respects. First, Reddit does not provide labeled sentiment, requiring the use of a text-based model like VADER to infer tone. Second, our analysis is focused on four hand-picked, politically exposed tech stocks rather than a wider market sample. This narrower scope allows us to examine behavior under politically charged conditions but may limit the generalizability of our findings.

The goal of this research design is not to establish causality but to detect whether patterns of alignment exist between political news, retail sentiment, and stock market activity. Focusing on Reddit as a platform provides access to a segment of investors who are reactive, vocal, and often emotionally engaged, making them particularly relevant for studying belief-driven responses to political narratives.

3.2 Data Collection

Sentiment data was collected from reddit using the Python Reddit API Wrapper (PRAW). The extraction process focused on English-language posts from major stock and investing subreddits, specifically r/stocks, r/investing, and r/wallstreetbets. Reposts were excluded to ensure data quality and reduce redundancy. For each company, posts were filtered based on relevant keywords matching each ticker, for example apple data looks for ‘AAPL' and/or ‘apple’, with the search being case-insensitive. Both the post titles and body texts were analyzed using the VADER sentiment analysis tool, which calculates a sentiment score for

each individual post. These scores were then aggregated on a daily basis to produce an overall sentiment score per day for each stock. The final dataset was stored in CSV format, containing daily average sentiment scores as well as a cumulative sentiment index, calculated recursively to reflect sentiment momentum over time.

Historical stock market data was retrieved using the Yahoo Finance API for the same study period, spanning from November 5, 2024, to March 13, 2025. For each company under analysis, daily data was extracted including open, close, high, low, and trading volume. This dataset provided the foundational market performance data necessary for comparison against sentiment trends, allowing for time-series alignment and analysis between retail sentiment and actual stock price movements.

3.2.1 Use of Data

Our method of capturing retail sentiment mirrors Cookson & Niessner (2020) in several respects. Like their use of StockTwits messages as a high-frequency sentiment source, we extract Reddit posts on a daily basis and compute an aggregate sentiment score per stock and day. While they benefit from user-labelled sentiment (bullish/bearish), our VADER-based approach still captures tone across thousands of posts, offering a close proxy. This methodological alignment supports the credibility of our sentiment data and allows us to test similar hypotheses, particularly how sentiment diverges or intensifies around politically salient events.

3.3 Data Processing

To enable comparative analysis, the sentiment data was combined with the corresponding stock market data by aligning observations on the shared ‘Date’ field. This ensured temporal consistency across datasets and allowed for a direct pairing of sentiment indicators with market performance metrics. As a result, each row in the combined dataset contained both the daily financial indicators, such as closing price, trading volume, and price ranges, and the VADER sentiment scores for that same calendar date. This integration was essential for enabling a meaningful analysis of potential relationships between retail sentiment and stock price movements.

To quantify daily stock performance, daily returns were calculated using the standard formula:

$$Return_t = \frac{Close_t - Open_t}{Open_t}$$

This metric offered a straightforward yet informative measure of how much a stock’s price changed during the course of a single trading day, relative to its opening value. By computing this return for each date in the dataset, we were able to capture the direction and magnitude of price fluctuations, which was crucial for investigating the extent to which daily sentiment may correlate with actual stock price changes.

3.4 Variable Selection

The variables selected for this study were chosen to comprehensively capture market performance, investor activity, and retail sentiment. Price variables included daily open, close, high and low values, offering a detailed view of each stocks' intraday behavior and volatility. In particular, the open and close prices were used to compute daily returns, serving as a core metric for analyzing price movements. Trading volume was incorporated to reflect daily investor activity, as changes in volume can signal shifts in market sentiment or reactions to external events, and often serve as a reinforcing can signal shifts in market sentiment or reactions to external events, and often serve as a reinforcing indicator of price significance. Finally, retail sentiment was quantified using VADER's compound sentiment score, averaged across all relevant Reddit posts per stock on a given day. This daily average captured short-term sentiment, while an additional cumulative sentiment score, calculated recursively, was included to represent longer-term sentiment momentum and broader trend in investor tone over time.

Table 2: Descriptive Statistics Table

a) AAPL

	Open	Close	Volume
mean	236.82	237.24	49678015.12
std	9.81	9.86	18724754.58
min	219.55	216.98	23234700.00
25%	228.22	228.09	39515600.00
50%	236.23	237.18	45118850.00
75%	244.58	244.70	54370350.00
max	257.91	258.74	147495300.00
count	86.00	86.00	86.00

b) MSFT

	Open	Close	Volume
mean	219.28	219.32	39267479.07
std	12.33	12.41	13935038.86
min	193.90	194.54	15007500.00
25%	208.54	208.45	29004525.00
50%	222.67	220.98	36662000.00
75%	228.84	228.96	43204825.00

max	239.02	242.06	88279200.00
count	85.00	86.00	86.00

c) NVDA

	Open	Close	Volume
mean	134.78	134.33	252292419.77
std	10.46	10.25	106589901.65
min	106.98	106.97	105157000.00
25%	129.97	129.94	188706350.00
50%	137.21	136.56	226860400.00
75%	141.29	140.23	282614550.00
max	153.02	149.42	818830900.00
count	86.00	86.00	86.00

d) TSLA

	Open	Close	Volume
mean	365.25	368.89	93080218.60
std	56.59	57.13	35217165.88
min	225.31	222.15	37167600.00
25%	336.59	331.12	68094525.00
50%	361.10	361.10	86418700.00
75%	410.56	406.08	112990525.00
max	475.90	479.86	210521600.00
count	86.00	86.00	86.00

3.5 Data Analysis

3.5.1. Rolling Correlation of Returns vs Sentiment

To capture the relationship between sentiment and stock performance, a rolling correlation analysis was conducted. For each trading day in the sample, the Pearson correlation coefficient was calculated between the daily returns and sentiment scores over the preceding 20-day window. This produced a time series of rolling correlation values, which was then visualized to identify any temporal patterns or shifts in the strength and direction of the relationship (see Figure 1a,b,c,d). The interpretation of this rolling correlation can be seen as follows; A positive and substantial correlation implies that higher levels of retail sentiment

tend to be associated with higher daily returns during the 20-day window in question. Conversely, a negative correlation suggests that sentiment and returns tend to move in opposite directions during that same period. This method allows for a dynamic understanding of how sentiment and market performance co-move, rather than assuming a static relationship across the entire study board.

3.5.2. Daily Returns and Sentiment Visualization

This analysis aimed to provide a direct visual and statistical comparison between daily stock returns and sentiment. The approach involved both time series, daily returns and corresponding sentiment scores, on the same graph across the full study period (see Figure 2a,b,c,d). This visualization allowed for an intuitive inspection of potential co-movement between the two variables. In addition to the visual analysis, a Pearson correlation coefficient was computed over the entire time series, accompanied by the corresponding p-value. This statistical measure served to evaluate the strength and direction of the linear relationship between sentiment and returns, as well as to test for its statistical significance. Interpretation was conducted both visually and numerically. By overlaying the two time series, one can assess whether peaks in sentiment tend to align with positive returns, and whether dips in sentiment coincide with market declines. The Pearson coefficient, along with its p-value, provided a concise summary of how closely the two series tracked each other throughout the observation window.

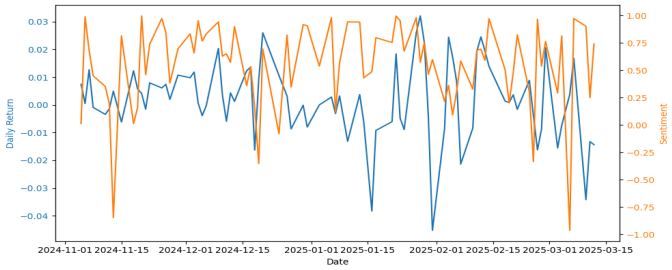
3.5.3. Volume, Sentiment, and Stock Changes

This analysis aimed to extend the investigation by incorporating trading volume into the relationship between daily returns and retail sentiment. The objective was to explore whether trading activity reflects or amplifies the influence of sentiment and price movement. The method involved the following steps for each stock in the dataset: Daily return was calculated using the daily return equation above, providing a measure of daily price movement. Daily volume change was computed as the percentage change in trading volume from the previous day, offering insights into fluctuations in investor activity. Average daily sentiment was extracted from Reddit-based VADER scores, reflecting the general tone of retail investor discussions. Using these three variables, a 3x3 Pearson correlation matrix was constructed for each stock. This matrix captured the pairwise linear relationship between sentiment, return, and volume change across the entire time period. Interpretation of the matrix is as follows: A positive correlation (closer to +1) between sentiment and returns suggests that higher sentiment is generally associated with higher daily returns. A strong correlation, either positive or negative, between volume and sentiment or volume and returns implies that trading activity is meaningfully linked to market perception or price movements. Conversely, correlation coefficients near zero indicate that the variables do not exhibit a consistent linear relationship over time.

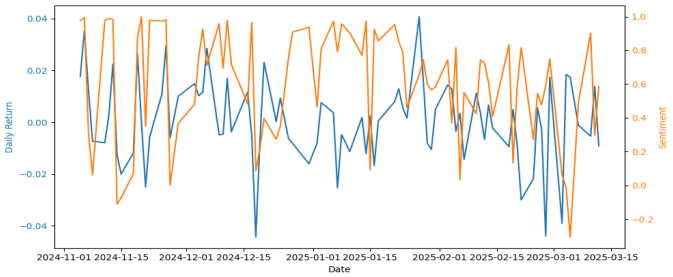
4. Results

4.1 Daily Return vs. Sentiment Time Series

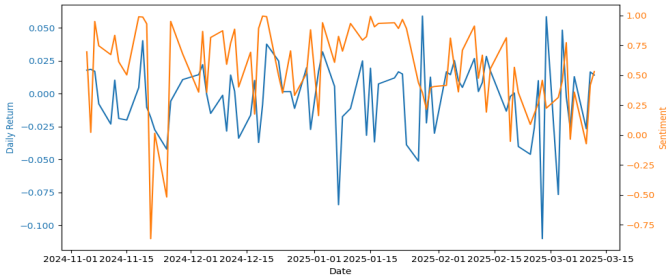
These graphs plot the daily VADER scores together with the daily returns of the four stocks analyzed.



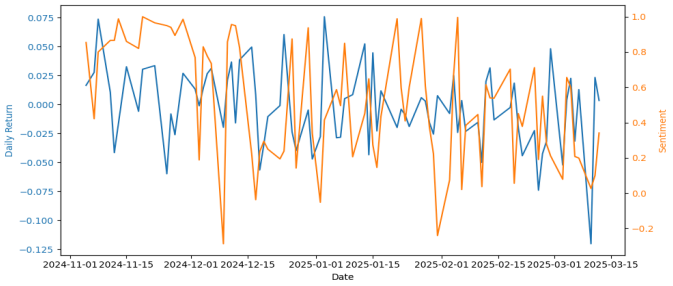
a) AAPL: Daily Returns and Sentiment Over Time



b) MSFT: Daily Returns and Sentiment Over Time



c) NVDA: Daily Returns and Sentiment Over Time



d) TSLA: Daily Returns and Sentiment Over Time

Figure 1: Timeplot of Daily Returns and Sentiment Score

Table 3: Pearson Correlation and p-value

	AAPL	MSFT	NVDA	TSLA
Pearson correlation	0.0141	0.1807	0.1619	0.2027
p-value	0.9003	0.1042	0.1436	0.0678

Apple displays noisy sentiment data with low return volatility, and there is no clear visual alignment between the two lines. The Pearson correlation is just 0.0141, with a p-value of 0.9003, confirming that there is no statistically significant relationship between sentiment and returns. This supports the view that Apple behaves more like a stable tech stock, largely unaffected by short-term shifts in retail sentiment.

Microsoft shows some periods where sentiment and returns visually align, particularly in the earlier part of the sample. However, this connection weakens over time. Numerically, it has a

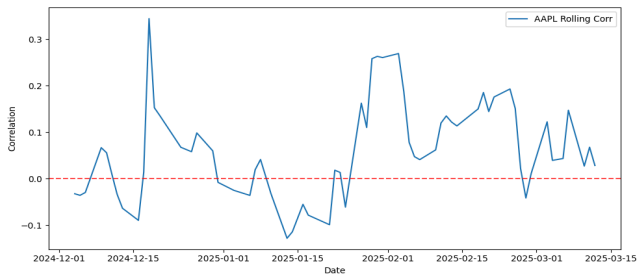
Pearson correlation of 0.1807 and a p-value of 0.1042, suggesting a weak, non-significant positive relationship. This complements the earlier rolling correlation graph, which shows an initial correlation that dissipates over time.

Nvidia's sentiment line remains relatively high, while returns are more volatile. There are moments of overlap, but overall the relationship is inconsistent. The Pearson correlation of 0.1619 with a p-value of 0.1436 indicates a weak, non-significant correlation. Still, this is higher than Apple's, hinting that Nvidia, being a front-runner in the AI space, may be more sentiment-driven, but not in a reliably predictive way.

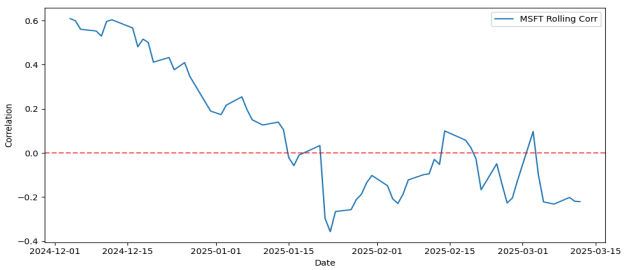
Tesla shows the most extreme fluctuations in both sentiment and returns, often with sharp movements that appear to coincide. It has the highest correlation among the four stocks: 0.2027, with a p-value of 0.0678. While this is still not statistically significant at the 5% level, it is the closest. This aligns with Tesla's high visibility in media and its polarized public perception, which may make it more reactive to sentiment shifts compared to the others.

4.2 Rolling Correlation Plots

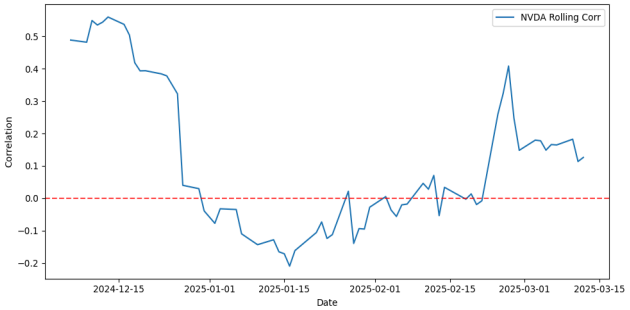
The following plots display the 20-day rolling Pearson correlation between daily Reddit sentiment scores and stock returns for each of the four companies under study. This approach enables a dynamic assessment of the sentiment-return relationship, revealing how it fluctuates over time rather than remaining constant. By tracking these short-term correlations, the analysis captures periods where sentiment appears more closely aligned with market performance, as well as intervals where the relationship weakens or potentially reverses. The patterns observed in each stock's correlation trajectory are discussed in the section below.

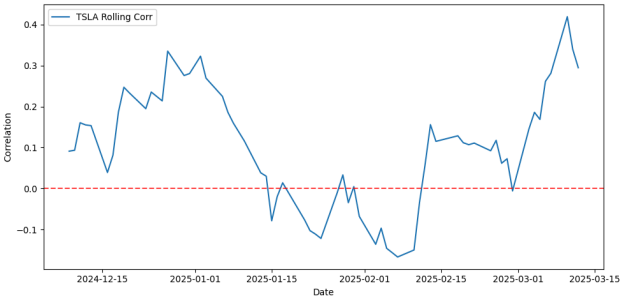


a) AAPL: 20-Day Rolling Correlation (Return vs. Sentiment).



b) MSFT: 20-Day Rolling Correlation (Return vs. Sentiment).





c) NVDA: 20-Day Rolling Correlation (Return vs. Sentiment). d) TSLA: 20-Day Rolling Correlation (Return vs. Sentiment).

Figure 2: 20 Day Rolling Correlation Plots

The AAPL rolling correlation lies mostly between -0.1 and +0.3. The changes show that the strength and direction of sentiment's correlation to returns for AAPL is mild and variable, sometimes modestly positive, sometimes near zero, and briefly negative.

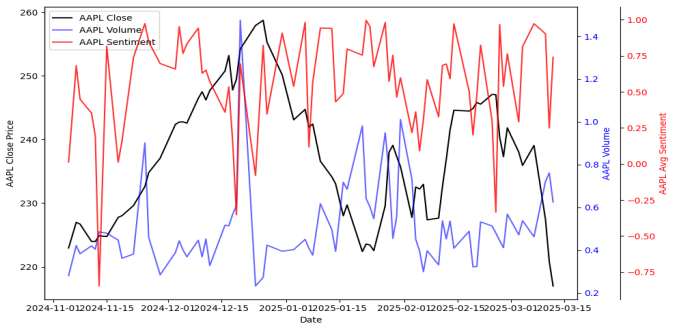
MSFT begins with a relatively high positive correlation (0.5-0.6 range) in the first few weeks. Over time, it trends downward towards zero and goes negative in late January. Through February and into March, the rolling correlation swings around zero, sometimes positive, sometimes negative. The takeaway is a fairly high correlation at the outset that dissipates over time ending in a mostly near-zero correlation.

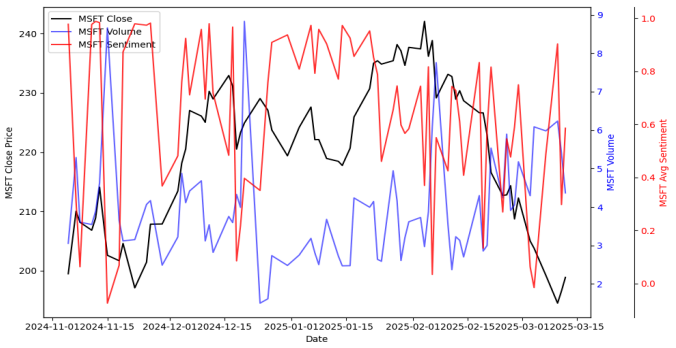
NVDA initially has a high daily positive correlation (around 0.5), then a steep drop below zero in January to around -0.2. Over February, the correlation climbed back towards positive territory and by early March spiked close to +0.4-0.5 before decreasing slightly toward mid-March. This indicates that sentiment-return relationships can flip, from positive to negative, over short windows and then rebound again.

TSLA starts slightly positive in late November, climbs up to about +0.35 around January then dips below zero in late January. In February, it rebounded sharply, going from slightly negative in early February all the way to around +0.4 or more by mid-March. Overall, TSLA shows significant swings in its short-term correlation, starting moderately positive, briefly becoming negative, then ending on a fairly strong positive note.

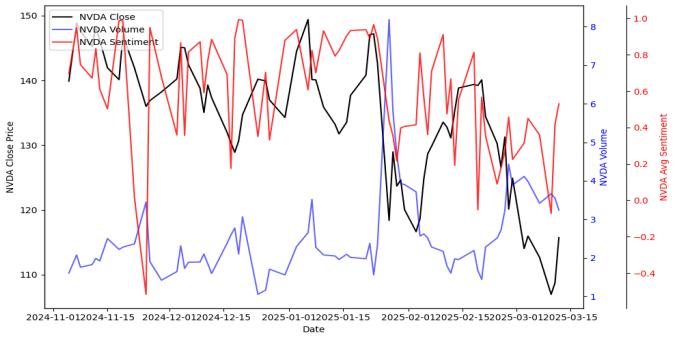
4.3 Trading Volume and Sentiment Correlation

The following graphs illustrate the daily development of trading volume, stock price, and average sentiment score for each of the four analyzed companies. These visualizations provide an overview of how sentiment and trading behavior evolve alongside market performance. Accompanying each graph is a correlation matrix summarizing the statistical relationship between the variables.

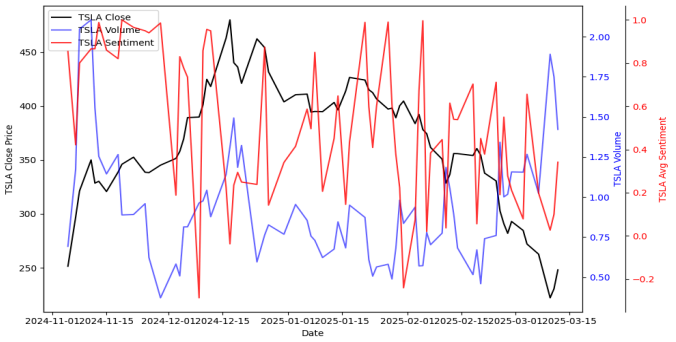




a) AAPL: Price vs. Volume Retail Sentiment



b) MSFT: Price vs. Volume Retail Sentiment



c) NVDA: Price vs. Volume Retail Sentiment

d) TSLA: Price vs. Volume Retail Sentiment

Figure 3: Trading Volume and Sentiment Correlation Plots

4.3.1 Correlation Matrices

Table 4: AAPL Correlation Matrix

	Return	Volume Change	avg Sentiment
Return	1.000000	-0.031339	-0.005908
Volume Change	-0.031339	1.000000	0.232296
avg Sentiment	0.005908	0.232296	1.000000

Between return and volume we see a near zero correlation, so Apple's daily price moves do not strongly depend on volume fluctuations. Between return and sentiment it was also near zero, implying little direct alignment between price changes and average Reddit sentiment. Between volume and sentiment we can see a modest positive link suggesting that when sentiment is higher, volume may tick up slightly.

Table 5: MSFT Correlation Matrix

	Return	Volume Change	avg Sentiment
Return	1.000000	-0.1083326	0.270175
Volume Change	-0.108326	1.000000	0.112534
avg Sentiment	0.270175	0.112534	1.000000

There is essentially no consistent pattern between price returns and volume changes. Mild positive relationship between return and sentiment, suggesting that on days with higher returns, sentiment is slightly more positive. Between volume and sentiment, the near zero indicates minimal connection.

Table 6: NVDA Correlation Matrix

	Return	Volume Change	avg Sentiment
Return	1.000000	-0.484070	0.226766
Volume Change	-0.484070	1.000000	-0.100817
avg Sentiment	0.226766	-0.100817	1.000000

Moderate negative correlation indicates that on days of bigger volume surges, the returns tend to be lower, or vice versa. In return vs sentiment a moderate positive correlation suggests some alignment between higher returns and more optimistic Reddit sentiment. Between volume and sentiment we have a nearly zero score, meaning that sentiment and volume are not clearly moving together.

Table 7: TSLA Correlation Matrix

	Return	Volume Change	avg Sentiment
Return	1.000000	-0.130664	0.146033
Volume Change	-0.120664	1.000000	-0.279015
avg Sentiment	0.146033	-0.279015	1.000000

Between return and volume there is a nearly zero correlation, so daily returns do not closely track volume changes. In return vs sentiment there is a slight positive correlation; no strong link but it still indicates a minor tendency for price gains to coincide with more positive sentiment. Lastly, in volume vs sentiment we can see a moderate negative relationship suggesting that higher volume changes are associated with somewhat lower sentiment, or possibly that increased trading activity corresponds to negative commentary.

4.4 Industry index

To assess whether the observed sentiment-return relationships are driven by stock-specific dynamics or broader sectoral trends, we compare the individual performance of the four selected companies to the Nasdaq-100 index, proxied through the QQQ ETF. This comparison helps isolate sentiment effects from general market movements by offering a benchmark for overall tech sector behavior during the study period.

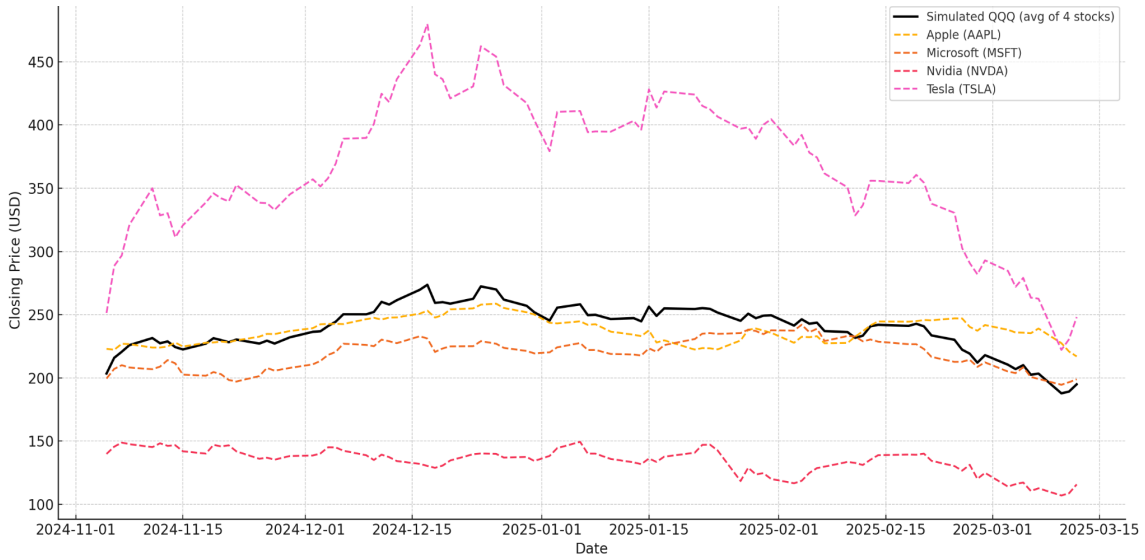


Figure 4: Simulated QQQ vs. Individual tech Stocks (Nov 5, 2024 - Mar 12, 2025).

While all four stocks broadly align with the overall trend of the simulated index, significant deviations are observed particularly for Tesla. TSLA displays pronounced volatility, with sharp price spikes in December followed by a steep decline through March. These movements depart notably from the more tempered trajectories of the other stocks.

5. Discussion

5.1 Analysis of Results

The data analytics revealed that the correlation between social media sentiment and daily stock returns is generally weak and statistically insignificant across the studied stocks. This finding suggests that while sentiment on Reddit may reflect investor mood, it has limited predictive power for daily stock price movements. These results are in line with the general findings of Semenova & Winkler (2021) and Lyócsa et al. (2021), which show that retail sentiment can amplify short-term price movements, but rarely constitutes a long-term explanatory factor. These results could be explained by the findings of Cookson & Niessner (2020), who show that retail investors tend to hold different and often conflicting beliefs, which leads them to disagree even when they are reacting to the same information. In our context, this suggests that the mixed sentiment on Reddit may reflect similar belief polarization, making it difficult for sentiment to consistently align with stock price movements.

The results also vary by stock and time period. Tesla and Nvidia exhibit the clearest variation, with periods where the correlation is positive, only to later turn negative or disappear entirely. One possible explanation is that Tesla and Nvidia have received significantly more media attention in recent years. Tesla, in particular, became a major topic of online hype a few years ago, emerging as one of the most talked-about stocks among younger investors (Thornhill, 2020). Elon Musk also became increasingly prominent online, appearing on major podcasts such as *The Joe Rogan Experience* and thus gaining global publicity. This could naturally

trigger a wide range of reactions, as public prominence often evokes both admiration and criticism. Furthermore, during the specific period that we have examined, Musk has faced substantial criticism (The Washington Post, 2025). While he may have gained even deeper admiration from his core group of ultra-fans, he has also attracted a significant amount of new hostility, so much so that a trend of vandalizing Teslas has emerged (Police 1, 2025). This polarization could naturally lead to highly varied reactions among current shareholders, retail investors, and institutional investors. It reinforces Cookson & Niessner (2020) theory that growing opinion diversity can intensify groupthink dynamics.

The upside is that this development supports the core argument of our reference article. However, the downside is that it becomes significantly more difficult to isolate and analyze the specific impact of retail sentiment on stock performance during this politically charged period. Since the recent news that have highlighted Musk's controversial behavior, including provocative gestures and appearing closely alongside President Trump, actions that have drawn sharp criticism from democratic voters, may explain why positive retail sentiment does not show a clear effect (Cole, 2025). Conversely, one might expect that a large volume of negative sentiment would correlate with price movements, but this, too, is not consistently observed. How can one truly determine what drives an equity's price? Universities teach about theories such as the EMH and fundamental investing principles from Graham, yet our own experience suggests these frameworks often fall short in practice (Dolan, 2024). There exists an ocean of unstructured behavioral effects that may influence prices in ways that are difficult to model or predict.

Nvidia is also one of the trendiest companies today, positioned at the forefront of the ongoing AI hype (Adebayo, 2025). The AI hype will naturally lead to a wide range of investor behaviors. For example, those who believe AI is a bubble will likely be against the hype around the company, choosing not to buy, possibly shorting the stock, or selling off their shares, which can put downward pressure on the price. On the other side, those who are excited by the hype will likely trade more actively, promote the stock, and push the price upward. To truly understand what drives the price, one would need to know every market participant's opinion and weigh them against each other in the data analysis, something that is probably impossible to do with today's tools.

Apple and Microsoft show more stable patterns, with generally low correlation between sentiment and returns. This makes sense, as these companies differ significantly from the other two stocks. Both have been around for around 50 years and are seen as more "stable tech companies," rather than media-driven or social media-hyped names constantly in the spotlight, whether positive or negative (Seeking Alpha, 2025). This supports the theory from Cookson & Niessner (2020) as well as our focus: political uncertainty, especially when political news directly involves a company, impacts retail sentiment in both positive and negative ways. This happens not only on Reddit but likely across other forums too. The equities that receive more media attention tend to be more volatile, since the different political happenings, such as Musk becoming a government official, boosts the confirmation bias of the people loving him versus hating him.

So while the results do not support the idea that retail sentiment has a clear and consistent effect on stock prices, they do support the view that political uncertainty leads to more divided opinions, which in turn increases volatility, especially in the most discussed stocks. However, one important limitation of this study is that we do not have data on how much of each stock is held by institutional investors versus retail investors. This information could significantly affect how much influence retail sentiment can actually have on price movements. We know that retail investors are more likely to act based on Reddit and other social media platforms, so the stocks with a higher share of retail ownership may react differently than those dominated by institutional holdings. If institutional investors hold positions worth billions while retail investors only hold millions, the impact of retail sentiment will naturally be more limited. We believe that mature companies like Apple and Microsoft likely have a higher proportion of institutional ownership, while a stock like Tesla may have more retail involvement and often gets labeled as a “meme stock.” This could explain why Tesla appears more sentiment-driven than the other stocks in our sample, and why it exhibits a notably high, and arguably overvalued, P/E ratio of 164 (Public, 2025).

Since our study differs in that sentiment is analyzed not only in isolation, but in the context of political uncertainty, we will now go deeper into analyzing the exact variation in the results in relation the the dates, to see what political news was in center at that specific time. Starting off with Apple, the correlation is mostly between -0.1 and +0.3. This means that the correlation is weak and variable, sometimes there is a small positive correlation (when Reddit becomes more positive, the stock goes up), sometimes no clear correlation at all, and sometimes a weak negative correlation (when Reddit becomes more positive, the stock goes down). In the graph, we see a peak in mid-December 2024, where the correlation reaches above +0.3. This is the highest point in the entire time series. While +0.3 is not considered a strong correlation statistically (e.g., +0.7 and up is strong), it is the clearest signal in the data that while Reddit posts about Apple were more positive, the stock price rose at the same time in that 20-day period. In mid-December 2024, an increase in Apple’s share price was noted. This increase coincided with positive news of strong sales forecasts for the iPhone 16 during the holiday shopping season. Additionally, the introduction of new AI-powered features, known as "Apple Intelligence," in the iPhone 15 Pro and iPhone 16 series was expected to increase consumer interest and drive the upgrade cycle (Kilgore, 2024). Overall, these factors could suggest that Apple’s stock rise in mid-December 2024 was mainly driven by strong sales forecasts and technological innovations rather than specific political events, or retail sentiment. At first, one might think it was due to Trump’s win on November 5, but it really does not cover the 20 day rolling window enough to be a credible reason.

Moving over to Microsoft, between November 25 and December 15, 2024, a relatively high but gradually decreasing correlation was observed between Reddit sentiment and Microsoft’s stock price. During this 20-day period, several significant events occurred that could potentially have impacted both the public’s perception of Microsoft and its stock price. First event is the FTC antitrust investigation into Microsoft (early December 2024). The U.S. The Federal Trade Commission (FTC) launched a comprehensive antitrust investigation into Microsof”s cloud computing, artificial intelligence, and cybersecurity businesses. This news

may have affected investor confidence and public discussions about the company (O'Brien, 2024). Next thing is information leak allegations on December 3 2024 (Feiner, 2024). Microsoft accused the FTC of leaking information about the ongoing antitrust investigation, which led to additional media attention and potentially impacted the company's image. On November 27, 2024, Microsoft announced that the launch of its Xbox game store on Android was delayed due to legal hurdles related to Google's alleged monopoly position. This event may have affected consumers' and investors' perception of Microsoft's position in the mobile gaming market (Lawler, 2024). While the Microsoft-specific events observed between late November and mid-December 2024 were not electoral in nature, they can still be classified as politically driven. The FTC's antitrust investigation, the related information leak, and regulatory barriers affecting Microsoft's Xbox store all stem from political or governmental processes. These developments reflect increased scrutiny from public authorities, and represent a form of political uncertainty. Such regulatory actions may dilute the predictive power of retail sentiment, as retail investors are less likely to anticipate complex legal or policy-driven outcomes. This could explain the observed decline in correlation between sentiment and returns during this period.

The correlation between Reddit sentiment and NVIDIA's stock price remained relatively high, hovering around +0.5, before declining sharply into January. Several key events during this period may help explain the observed fluctuations in sentiment and stock returns. On December 9, 2024, Chinese regulators (SAMR) launched an antitrust investigation into NVIDIA's $6.9 billion acquisition of Mellanox Technologies, originally completed in 2020. The investigation focused on whether NVIDIA had violated Chinese anti-monopoly conditions tied to the original approval of the deal. This announcement introduced uncertainty around NVIDIA's operations in China, leading to a 2% stock drop and likely triggering negative sentiment among retail investors on forums like Reddit (Mcdade, 2024). Even though we might not confirm in a statistically significant way that the correlation exists, we can see clear qualitative signs of political news leading to stock price surges. On the other hand, it could be argued that even though the most hyped company online right in the most hyped AI bubble, is not correlated to the stock prices due to the effect of the negative political news, arguing for the significant political news.

The correlation between Reddit sentiment and Tesla's stock price remained positive but volatile, fluctuating in the +0.1 to +0.3 range. One significant development during this period was Tesla's announcement that it would delay the construction of a $10 billion gigafactory in Monterrey, Mexico. This decision followed proposed 25% import tariffs from the incoming Trump administration (Revell, 2024). The event likely affected investor sentiment and online discussion, illustrating how political uncertainty can influence both public perception and short-term price dynamics.

To conclude, we have data analysis on a 20-rolling day basis that confirms a weak correlation, with variable results. We can also see a lot of effects on how political news affects the stock prices, this supports and extends the conclusions of Cookson & Niessner (2020). Similar to their observation of persistent belief divergence in social trading networks,

we observe that Reddit sentiment around politically charged events is often contradictory and highly polarized. For instance, Tesla-related sentiment surged both positively and negatively following Trump's tariff threats, reflecting how political narratives activate different investor subgroups. This echoes the "belief reinforcement" mechanisms Cookson & Niessner document, now shown to be activated by political, rather than purely financial, stimuli. This conclusion is also supported by Pastor & Veronesi (2012) and Baker et al. (2016), who have shown that political uncertainty increases market volatility. Our results suggest that these political cycles act as reinforcers of retail sentiment, rather than independent influencing factors. During periods of higher rhetoric from Donald Trump (such as in January and February 2025), we also see temporary jumps in the correlation between sentiment and returns, especially in stocks with strong retail links like TSLA and NVDA.

5.1.2 Correlation Between Sentiment and Stock Performance day by day

Besides the rolling analysis, we also looked at how sentiment and returns moved together over the full time period. The table shows the Pearson correlations. Tesla had the highest correlation at +0.2027, followed by Microsoft (+0.1807) and Nvidia (+0.1619). But none of these are statistically significant, since all the p-values are above 0.05. This means that Reddit sentiment did not clearly predict stock returns across the full period from November 2024 to March 2025. Still, the higher numbers for Tesla and Nvidia match what we saw earlier, some stocks with a strong online presence seem to have a closer link between sentiment and price, at least for certain periods. This supports earlier research by Semenova & Winkler (2021), who found that social media can move prices short-term, but not over longer time frames.

If we look at the time series plots, we can see that Tesla and Nvidia sometimes show similar movements in sentiment and returns, especially when political news is in focus. For example, Tesla had a spike in positive sentiment in early December, which also matched with a short-term price increase. But other times, like in February, sentiment was high without any price reaction. This up-and-down pattern reflects the idea from Cookson & Niessner (2020), who say that social media does not always help investors agree. Instead, people mostly follow others who already think like them. So the same political news can lead to very different opinions, some Reddit users think Trump's policies are good for business, while others do not.

Apple and Microsoft had lower and more stable correlations (AAPL: +0.0141, MSFT: +0.1807), and their p-values were much higher. This suggests that these older, more stable tech companies are less affected by retail sentiment and more influenced by traditional business fundamentals.

In summary, our results support the ideas from Cookson & Niessner (2020). Their findings demonstrated that social trading platforms do not foster consensus among users; instead, they reinforce existing beliefs, causing investors to maintain their pre-existing views. Our analysis reflects a similar pattern. Even during big political events, there is no clear pattern between sentiment and stock returns. Instead, we see different groups reacting differently, which

makes sentiment harder to use as a signal. Political events, especially when they involve Trump, seem to amplify emotional reactions rather than give clear market direction. This matches findings by Pastor & Veronesi (2012) and Baker et al. (2016), who found that political uncertainty often increases market volatility and makes it harder to read investor sentiment.

5.1.3 Sentiment and Volume: Behavioral Response Rather Than Price Impact

In this part of the study, we looked at how Reddit sentiment connects not only to price changes, but also to trading volume. The goal was to see if positive or negative sentiment leads to more trading activity, even when prices do not move much. Looking at the correlation matrices, we see that price returns and volume changes are mostly uncorrelated across all four stocks. This means that daily price movements do not depend strongly on how much the stock was traded that day. Instead, it is the connection between sentiment and volume that shows more interesting results. For Apple, sentiment and volume had a small positive correlation (+0.22), meaning that on days when Reddit sentiment was more positive, trading volume was slightly higher. The connection to returns, however, was weak (+0.04). This shows that sentiment may activate trading, but does not clearly affect the price. For Microsoft, the volume-sentiment correlation was even weaker (+0.05), and the return-volume link was near zero. Again, this means that when sentiment shifts, investors may become more active, but it does not consistently lead to higher or lower prices. Nvidia showed a special pattern. There was a moderate positive link between sentiment and returns (+0.26), but almost no connection between sentiment and volume (-0.08). This could mean that Reddit sentiment sometimes predicts price changes for Nvidia, but not because more people are trading, possibly due to more concentrated buying from certain groups. The clearest volume-sentiment effect came from Tesla. Here, sentiment and volume had a moderate negative correlation (-0.29). This means that on days with high trading volume, Reddit sentiment tended to be more negative. One possible explanation is that controversial news about Elon Musk or political issues made people more active in trading, both retail and institutional, while the online discussions turned more critical.

These results are in line with the idea that Reddit acts as a coordination platform, where people are activated by strong opinions, but not always in a clear direction. At times, heightened sentiment triggers increased trading activity, at other times, it has little to no effect. This is similar to what Da, Engelberg & Gao (2011) and Tetlock (2007) found: social media sentiment may not drive prices directly, but it often drives behavior, especially among retail traders. Finally, this supports the idea from Cookson & Niessner (2020) that retail investors respond to sentiment in different ways, some react strongly to political or emotional news, while others remain unaffected. This creates mixed market signals, especially in politically uncertain times like the 2024 election. The result is more trading, but not always price movement, which shows how social media can increase market noise without adding clear information.

5.1.4 Comparison to Industry Index

In order to test whether the sentiment effects observed in our study are truly stock-specific or merely reflections of broader industry trends, we compare our individual stock data to the QQQ ETF, a proxy for the tech sector as a whole. This analysis allows us to assess whether movements in our four selected stocks are unique or simply mirror general market behavior. The average performance of Apple, Microsoft, Nvidia, and Tesla (see Figure 4) tracks the actual QQQ index relatively closely over the study period. This suggests that on aggregate, the four stocks broadly reflect sector-wide developments. However, when analyzing each stock individually, deviations become more apparent. Tesla, in particular, displays large price swings that do not align with the broader index. These movements appear to be more sentiment-driven and idiosyncratic, as supported by the correlation analysis discussed in previous sections. Nvidia also shows moderate divergence at certain points, particularly during AI-related news cycles or China-specific political developments. Apple and Microsoft, by contrast, track the industry index more closely, indicating that their price movements may be less influenced by retail sentiment and more driven by overall market forces. These findings support our earlier conclusion that sentiment effects are more visible in retail-heavy, high-attention stocks like Tesla and Nvidia, where public narratives and political news drive independent movement. In contrast, Apple and Microsoft appear more insulated from such sentiment dynamics, behaving more like benchmark stocks that follow the broader index.

A possible reason for this difference is the underlying market positions of the companies. Apple and Microsoft operate in more mature and consolidated markets, where they hold near-monopoly positions in the Western world. Apple's main competitor is Samsung, which has also been an established player for decades (Oberlo, 2023). Similarly, Microsoft dominates across personal computing and enterprise software (Sharjeel, 2025). In contrast, Tesla and Nvidia operate in sectors that are still developing rapidly. Electric vehicles are a relatively new market, and competition has grown only recently. Tesla stock, for example, dropped sharply after BYD announced new lower-cost EV models, and similar market reactions followed Audi's earlier announcement to fully electrify its vehicle fleet (Kolodny, 2025). For Nvidia, dominance in AI chips has been strong, but sentiment has quickly turned when news of Chinese alternatives surfaced (Lin, 2025). These are novel firms with less entrenched competitive moats, which could explain why they react more strongly to sentiment shifts and news cycles compared to the more "established" tech giants.

5.2 Quality of Study

Following the steps of Cookson & Niessner (2020):

Importantly, our empirical framework draws inspiration from Cookson & Niessner (2020), whose study is published in a leading finance journal and demonstrates that investor disagreement can be meaningfully measured using social media data. By structuring our sentiment and trading activity analysis in a similar way, though with Reddit/VADER rather than StockTwits/self-labelling, we ensure that our study rests on a well-established methodological foundation. This also sets the stage for future refinements, such as

distinguishing between different subgroups of Reddit users, mimicking their decomposition into within-group vs. cross-group disagreement.

Methodological Considerations:
Correlation is a useful tool for identifying correlations, but it does not prove causation. A sentiment surge may coincide with a price movement, but we can not say with certainty what came first without including time lag or event studies, which may be developed in future research.

Transparency:
All code for data collection, sentiment analysis, and analysis is documented. By linking political events to retail-driven behavior in the market, we hope to open up for continued research in behavioral finance and microstructured market studies.

Lastly, our study is transparent and replicable. The data sources are publicly available, and our code is well-documented, which allows future researchers to validate and expand upon our results. While generalizability is limited to the selected stocks and time period, we believe the methodological framework can be applied across sectors and geographies.

6. Summary

This study examines how retail investor sentiment on Reddit influences the market behavior of four major tech stocks (Tesla, Nvidia, Apple, and Microsoft) during a period of heightened political uncertainty. By combining sentiment analysis of Reddit posts with stock price and trading volume data from November 2024 to March 2025, we investigate whether retail-driven sentiment aligns with financial market outcomes, and how these relationships evolve during politically charged periods.

Our results show that the correlation between sentiment and daily returns is generally weak, though it varies by stock and over time. Tesla and Nvidia display the clearest sentiment-driven fluctuations, with short-term windows where sentiment and price move in the same direction, followed by periods of decoupling or even inverse correlation. In contrast, Apple and Microsoft show more stable and weaker patterns, suggesting that larger, more mature tech companies are less reactive to social sentiment.

We find a stronger and more consistent relationship between sentiment and trading volume, particularly for Tesla. This indicates that Reddit sentiment may act more as a behavioral trigger, prompting investor activity, rather than a reliable signal for price changes. This supports the idea that retail forums serve more as coordination platforms than as sources of price-relevant information.

Overall, our findings reinforce prior research on short-term speculation and behavior-driven trading (e.g., Da, Engelberg & Gao, 2011), while also extending the framework of Cookson

& Niessner (2020) by showing how belief divergence and emotional responses among retail investors can intensify during political uncertainty. In this context, social media becomes a channel where political narratives and market reactions blend, creating mixed and noisy signals, especially in retail-heavy stocks. This study contributes by highlighting how sentiment-driven behavior interacts with macro-political cycles, not just market fundamentals.

From a deeper look into our main article, Cookson & Niessner (2020), who analyze different beliefs among retail investors, based on investment styles that themselves registered, in a politically neutral setting, our study introduces the role of political uncertainty. While they show that cross-group disagreement is persistent and not driven by gradual information diffusion, we observe that political events, especially during this timeframe surrounding Donald Trump's candidacy, increase sentiment volatility and reinforce emotional disagreements in retail communities.

Furthermore, unlike their broader dataset of many firms, we intentionally narrow our focus to four high-attention tech stocks that are both popular on Reddit and politically sensitive. This allowed us to zoom in on how sentiment behaves in the stocks where disagreement is most likely to materialize, to actually be able to know the political news effects. While this limits generalizability, it reflects our deliberate strategy to test sentiment mechanisms under our focused conditions.

As well, our contribution to their study and this topic in general, lies in showing that retail belief divergence does not just persist but also intensifies when political narratives are all around the news. This insight matters because it suggests that political cycles, such as elections, can temporarily increase the behavioral noise in financial markets. This complements Cookson & Niessner findings and provides a new lens for understanding how investor disagreement evolves when macro and micro narratives merge.

The final step, comparing the stock movements of our selected companies to the Nasdaq-100 index (QQQ), we confirm that while the average trajectory of all four stocks aligns with broader market trends, certain stocks, particularly Tesla and Nvidia, deviate more substantially. This supports our claim that sentiment-driven volatility is not just a reflection of sector-wide shifts but also a result of stock-specific narratives amplified by political news and social media. This strengthens the study's contribution by showing that belief divergence and sentiment-driven effects are more pronounced in politically exposed, retail-dominated stocks than in tech benchmarks generally.

6.1 Limitations

This study has several methodological limitations. First, the time period analyzed (November 2024 to March 2025) is politically interesting but relatively short, which limits our ability to draw conclusions about long-term effects or broader market cycles. Second, our sample

includes only four large-cap tech stocks, which reduces the generalizability of our findings to other sectors, such as energy, financials, or small-cap stocks

In terms of sentiment analysis, we have relied on VADER. While it performs well on large datasets, it may miss nuances such as sarcasm, slang, or political irony, language characteristics that are especially common on Reddit. However, VADER, as a rule based model, lacks the ability to understand deeper semantic meaning or evolving linguistic conventions that emerge in real-time communities such as Reddit. Additionally, it treats each post as a standalone text, ignoring thread-level or temporal context that might influence sentiment interpretation. Furthermore, we only included English-language posts, which may introduce a bias, particularly since many Reddit users write in English even if they are international. This language filter could skew the sentiment sample and underrepresent the views of non-English-speaking communities. Additionally, one critical issue is VADER's inability to accurately distinguish between contextually similar terms, which can lead to misclassifications. For instance, in the case of Apple Inc, VADER may misinterpret references to Apple as the fruit rather the technology company, resulting in erroneous sentiment scoring. This semantic ambiguity can affect the reliability of sentiment analysis, especially for companies with commonly used names. Given that no specific filtering techniques were applied to resolve this issue, the risk of misclassification remains, potentially impacting the accuracy of the sentiment scores.

Additionally, our sentiment data source (Reddit) itself represents a limitation. Reddit users are not necessarily representative of all retail investors; they tend to be younger, more speculative, and more concentrated in certain market segments like technology and cryptocurrencies. This means our results capture the behavior of a specific, highly vocal subset of the retail market rather than the retail population as a whole. Furthermore, the internal dynamics of Reddit data may introduce additional biases. Specifically, high-volume users, who post frequently, alongside automated accounts can exert a disproportionate influence on sentiment analysis, potentially distorting the measured sentiment and leading to misleading conclusions. This concentration of activity among a small subset of users can cause sentiment singlas to reflect the views of a hyper-engaged minority rather than amore balanced cross-section of retail investors.

Despite these limitations, the study provides a structured and transparent methodology, and we believe the findings are valuable within the context of tech stocks, political uncertainty, and social-media-driven sentiment.

6.2 Suggestions for further research

Future studies should consider a longer time span and include a broader range of sectors beyond large-cap tech stocks to enhance the generalizability of findings. For instance, examining sentiment-driven market reactions across different industries, such as Big Tech (e.g., Microsoft, Apple), Big Banks (e.g., JPMorgan, Goldamn Sachs), Telecoom (e.g., AT&T, Verizon), and Defense (e.g., Lockheed Martin, Raytheon), could reveal sector specific

pattern in how social media sentiment impact stock performance. Such a comparative approach may help identify which industries are most susceptible to sentiment-driven volatility and how these effects differ depending on industry characteristics, such as regulatory sensitivity or public visibility. Furthermore, an important extension would be to test for lagged effects by analyzing whether Reddit sentiment influences market behavior with a 1-2 day delay. This would provide insights into the speed of sentiment transmission from social media to actual stock prices.

In addition, conducting event studies around specific political developments, such as Trump's campaign speeches, policy announcements, or geopolitical events, could clarify the causal link between political events, sentiment shifts, and market reactions. A second area for future research is to further explore the behavior of institutional investors, particularly through the lens of short interest data in response to sentiment peaks. This can provide valuable insights into how professional investors react to sentiment-driven retail flows, especially during periods of heightened political uncertainty. Such analysis could reveal whether institutions take contrarian positions when retail sentiment is strongly bullish or bearish.
Finally, further methodological development is encouraged. This includes collecting sentiment data from additional sources (e.g. X, financial news sites) and applying more advanced natural language processing (NLP) models beyond VADER. Leveraging transformer-based models could capture more nuanced investor markets. These improvements would allow future research to provide a more comprehensive view of how sentiment influences asset prices, not only during speculative bubbles but also in response to real-world macro and political events.

7. References

Adebayo, K.S. 2025, “Experts Predict the Bubble May Burst for AI in 2025”, *Forbes*.

Antweiler, W. & Frank, M.Z. 2004, “Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards”, *The Journal of Finance*, Vol. 59(3), pp. 1259–1294.

Assis, C. 2024, “Apple to see strong Christmas iPhone sales that launch a 'renaissance of growth'”, *MarketWatch.*

Baker, S.R., Bloom, N. & Davis, S.J. 2016, “Measuring Economic Policy Uncertainty”, *National Bureau of Economic Research*, NBER Working Paper No. 16128.

Cookson, J.A. & Niessner, M. 2020, “Why Don’t We Agree? Evidence from a Social Network of Investors”, *The Journal of Finance*, Vol. 75(1), pp. 173-228.

Da, Z., Engelberg, J. & Gao, P. 2011, “In Search of Attention”, *The Journal of Finance*, vol. 66, no. 5, pp. 1461–1499.

Dolan, B. 2024, “Benjamin Graham's Timeless Investment Principles”, *Investopedia*.

Einhorn, N., Fisch, J.E., Gramitto Ricci, S.A., Le, M. & Sautter, C.M. 2023, “The Retail Investor Report”, *Public*.

Fama, E.F. 1970, “Efficient Capital Markets: A Review of Theory and Empirical Work”, *The Journal of Finance*, Vol. 25(2), pp. 383–417.

Feiner, L. 2024, “Microsoft accuses FTC of leaking news of its antitrust investigation”, *The Verge*.

Financial Industry Regulatory Authority (FINRA). 2021, “Following the Crowd: Investing and Social Media”, *FINRA*.

Financial Times. 2020, “Tesla and the audacity of hype”, *Financial Times*.

Hutto, C.J. & Gilbert, E. 2015, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text”, *Georgia Institute of Technology*, Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM-2014).

Investopedia. 2024, “Fundamental Analysis: Principles, Types, and How to Use It”, *Investopedia*.

Kolodny, L. 2025, “Tesla drops 6% after BYD partners with DeepSeek, Musk adds to DOGE distractions with OpenAI bid”, *NBC New York*.

Lawler, R. 2024, “Microsoft says it built an Xbox game store on Android but can’t launch it”, *The Verge*.

Lin, L. 2025, “China’s Huawei Develops New AI Chip, Seeking to Match Nvidia”, *The Wall Street Journal*.

Lyócsa, Š., Molnár, P. & Výrost, T. 2021, “Stock market volatility forecasting: Do we need high-frequency data?”, *International Journal of Forecasting*, Vol. 37(4), pp. 1092–1110.

Malhotra, R. & Jones, C. 2025, “Reddit Data in Quantitative Financial Models: Evolution and Implications Post GameStop and AMC Short Squeeze”, *Unpublished Thesis*.
McDade, A. 2024, ‘Nvidia Stock Slips as Chinese Regulators Announce Anti-Monopoly Probe’, *Investopedia*.

McElhaney, A. 2023, “Institutional Investors Are Using Reddit to Make Investment Decisions. Here’s Where They Lurk”, *Institutional Investor*.

Mercanti, L. 2024, “Natural Language Processing (NLP) in Finance: How AI is Transforming Market Analysis”, *Medium*.

Moody’s Analytics. 2024, “The Power of News Sentiment in Modern Financial Analysis”, *Moody’s*.

Münster, M., Reichenbach, F. & Walther, M. 2024, “Robinhood, Reddit, and the News: The Impact of Traditional and Social Media on Retail Investor Trading”, *Journal of Financial Markets*, Vol. 71, p. 100929.

Newsweek. 2025, “Would Elon Musk's Gesture Be Deemed a 'Nazi Salute' in Germany?”, *Newsweek*.

Nugent, J. 2022, “History of Wall Street Bets: From the Founding to 2023”, *Market Rebellion*.

Oberlo, 2024, “US Smartphone Market Share”, *Oberlo*.

O’Brien, M. 2024, “FTC opens Microsoft antitrust investigation that Trump administration must carry on or drop”, *AP News*.

Pano, T. & Kashef, R. 2020, “A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19”, *Big Data and Cognitive Computing*, Vol. 4(4), p. 33.

Pastor, L. & Veronesi, P. 2010, “Uncertainty About Government Policy and Stock Prices”, NBER Working Paper No. 16128, *National Bureau of Economic Research*.

Police 1. 2025, “‘No repercussions to lawlessness’: Colo. Tesla dealership vandalized for 5th time in 2025”, *Police 1*.

Preece, R. 2024, “The Finfluencer Appeal: Investing in the Age of Social Media”, *CFA Institute*.

Public. 2025, “Tesla (TSLA) P/E Ratio: Current & Historical Analysis”, *Public*.

Reddit. 2024, “Reddit's Top Mentioned Stocks of 2023 are up 70% This Year”, *Reddit*.

Reichenbach, F. & Walther, M. 2023, “Financial Recommendations on Reddit, Stock Returns and Cumulative Prospect Theory”, *Digital Finance*, Vol. 5, pp. 421–448.

Revell, E. 2024, “Musk says Tesla's Mexico factory on pause over Trump's tariff pledge“, *Fox Business*.

Salvucci, J. 2023, “A timeline of the GameStop short squeeze”, *TheStreet*.

Seeking Alpha. 2025, “Apple Is Seeking A Catalyst, Like Ballmer's Microsoft”, *Seeking Alpha*.

Sharjeel, M. 2025, “The Top 10 Biggest Software and Tech Companies in the World”, *The Designs Firm*.

Shefrin, H. 2024, “How Overvalued Is Tesla’s Stock, Really?”. *Forbes*.

Semenova, V. & Winkler, J. 2021, “Reddit’s self-organised bull runs: Social contagion and asset prices”, *University of Oxford,* MPRA Paper No. 107575.

Sugihartono, S. 2024, “The Power of Social Media to Influence Political Views and Geopolitical Issues: TikTok, X and Instagram”, *Modern Diplomacy*.

The Washington Post. 2025, “Elon Musk stirs controversy over hand gesture at Trump rally”, *The Washington Post*.

Vantage Markets. 2024, “The Psychology of Market Reaction: Understanding Investor Behaviour”, *Vantage Markets*.

7. Appendix

7.1 Defining Variables

Table 8: Definitions of Dependent and Independent Variables

Variable	Type	Definition	Source
Daily Return	Dependent	Calculated as $\frac{Close-Open}{Open}$; measures the intraday percentage change in stock price.	Yahoo Finance
Avg Sentiment Score	Independent	Daily average of VADER compound scores from Reddit posts per stock, capturing short-term sentiment.	Reddit API
Cumulative Sentiment Score	Independent	Recursively calculated index of past average sentiment values, capturing sentiment momentum.	Reddit API
Open Price	Control	The price at which the stock opened on a given trading day.	Yahoo Finance
Close Price	Control	The price at which the stock closed on a given trading day.	Yahoo Finance
High Price	Control	The highest price reached by the stock during the trading day.	Yahoo Finance
Low Price	Control	The lowest price reached by the stock during the trading day.	Yahoo Finance
Trading Volume	Control	The total number of shares traded during the day; reflects market activity and investor interest.	Yahoo Finance

7.2 Robustness Check

In order to check the robustness of our study, we have conducted a permutation test. To test whether the relationship between sentiment and the outcome variable is genuine or simply due to chance, we randomly shuffle the sentiment values and rerun the same regression multiple times. This process creates a distribution of estimated coefficients (β) that would be expected if there were no actual relationship, essentially a null distribution. By comparing the original β from the unshuffled data to this distribution, we can assess whether the observed effect lies in the extreme ends (suggesting a real effect) or near the center, indicating it might just be random variation.

Model specification

The analysis uses a basic Ordinary Least Squares (OLS) regression model applied to individual stocks:

$$Return_t = \alpha + \beta\, x\, Sentiment_t + \epsilon_t$$

Table 9: Description of Variables Used in the Permutation-Based OLS Regression

$Return_t$	Represents the daily return of a stock, calculated as the percentage change in its closing price from the previous day: $\frac{Close_t - Close_{t-1}}{Close_{t-1}}$
$Sentiment_t$	The explanatory variable, capturing the average compound sentiment score (based on VADER) for day t.
α	Denotes the intercept, or the expected return when sentiment is neutral (i.e., equal to zero).
β	is the key coefficient of interest, reflecting the average change in return associated with a one-unit change in sentiment.
ϵ_t	The error term, accounting for any variation in returns not explained by sentiment.

Interpreting β

A β value of 0.002 implies that a one-point increase ini sentiment is associated with an average increase of 0.2% in the stock's return on the following day. If β is negative, it suggests that more positive sentiment is linked to lower returns, indicating an inverse relationship between sentiment and return.

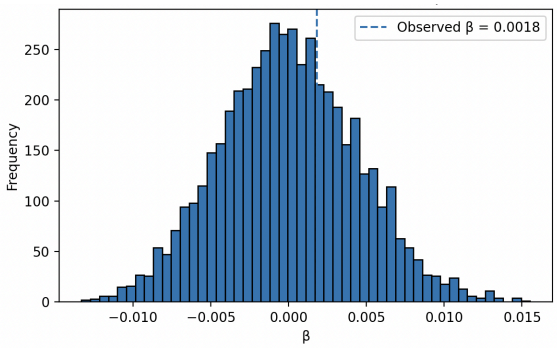


Figure x: AAPL Permutation Distribution of β

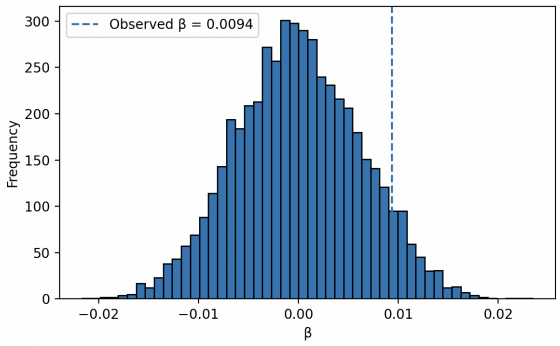


Figure x: MSFT Permutation Distribution of β

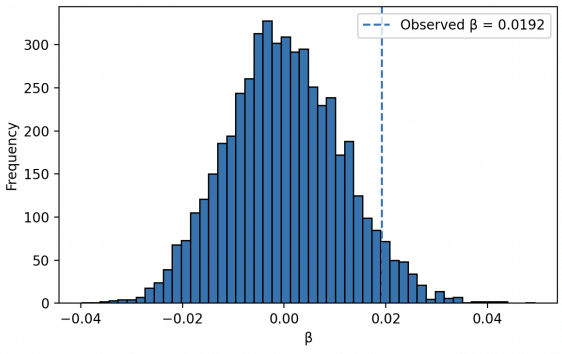


Figure x: VDA Permutation Distribution of β

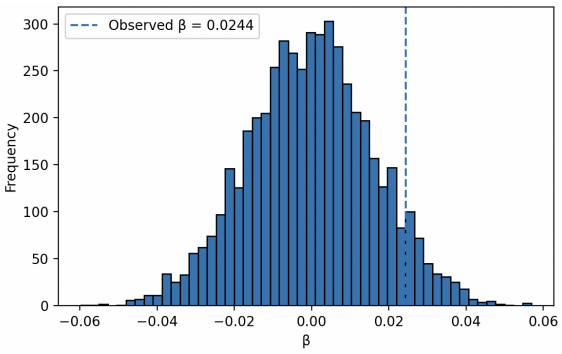


Figure x: TSLA Permutation Distribution of β

Table 10: Permutation test p-value

	AAPL	MSFT	NVDA	TSLA
p-value	0.6756	0.1368	0.0962	0.1358

Apple's observed beta of 0.0018 is very close to the center of the null distribution, suggesting a very weak positive association between sentiment and returns. The high p-value of 0.6756 confirms that this result is not statistically significant. The observed effect falls within what could occur by random chance; thus, no meaningful relationship is detected. Microsoft's beta of 0.0094 is more positive and lies farther into the right of the null distribution. This implies a somewhat stronger link between positive sentiment and next-day returns. However, the p-value of 0.1368 remains above conventional significance thresholds, indicating that while there may be a trend, the result is not statistically significant. Nvidia shows a relatively high beta of 0.0192, suggesting a stronger positive effect of sentiment on returns. The p-value of 0.0962 is the lowest among the four and borders on significance at the 10% level. This points to a potentially meaningful relationship, though it remains inconclusive under stricter standards, such as the 5% thresholds. Tesla has the highest observed beta of 0.0244, indicating a clear directional association between sentiment and returns. However, the p-value of 0.1358 suggests that this result also lacks statistical significance. Despite the large effect size, we cannot rule out the possibility that it occurred by chance, given the structure of the null distribution.

7.3 AI Transparency

Following SSE's guidelines, we utilized generative AI, specifically ChatGPT 4.0, as a supportive tool during the early stages of our research. We primarily used it for inspiration and brainstorming, helping us explore different perspectives and refine our ideas. However, AI was not used as a basis for analysis or as a source of content in our thesis. Additionally, we occasionally relied on AI to enhance the clarity and readability of our writing by checking grammar, spelling, sentence structure, and flow. Despite this, all AI-assisted suggestions were critically reviewed and adapted to ensure full ownership of our work. To maintain academic integrity and accuracy, we deliberately avoided incorporating AI-generated text directly into the thesis.