

Stock Data Scraping and Sentiment Analysis Project

- By Asawari Hire

OBJECTIVE

1. Business Objective

Enhance investment decision-making through social media sentiment analysis.

- Provide real-time insights on market sentiment for portfolio companies.
- Identify positive or negative trends in public discussions to anticipate stock movements.
- Assist investors and analysts in making data-driven decisions by leveraging X (formerly Twitter) sentiment.
- Improve risk assessment by detecting potential volatility based on sentiment shifts.

2. Technology Objective

Develop a scalable and automated sentiment analysis pipeline.

- Implement web scraping with Selenium to collect live tweets on selected stock tickers.
- Utilize VADER sentiment analysis to classify tweets as positive, negative, or neutral.
- Optimize performance with automated scrolling, randomized delays, and retries to maximize data accuracy.
- Store results in structured CSV files for further analysis and visualization.
- Ensure adaptability for future integration with AI-based trading models.

PROJECT OVERVIEW

This project focuses on developing an automated sentiment analysis pipeline to track and analyze social media discussions about specific stock tickers on X (formerly Twitter). The system aims to provide real-time market sentiment insights to assist investors and analysts in making data-driven investment decisions.

Using Selenium-based web scraping, the system collects live tweets mentioning predefined stock tickers such as RCAT, DKNK, and LYFT. The collected data is processed using the VADER sentiment analysis model, which classifies each tweet as positive, negative, or neutral. This sentiment information helps identify market trends, anticipate stock movements, and assess potential volatility based on public discussions.

The project output consists of structured CSV files containing stock symbols, tweet text, and sentiment classifications, which can be further analyzed and visualized. Additionally, the system is designed for scalability and adaptability, ensuring future integration with AI-based trading models or advanced predictive analytics.

INPUT DATASET

1. **Stock Tickers**
 - A predefined list of stock symbols (e.g., RCAT, DKNNG, LYFT, etc.) to track discussions on X (formerly Twitter).
2. **Live Tweets**
 - Scraped tweets containing mentions of stock tickers in the format "\$TICKER" (e.g., "\$RCAT is set to rise").
 - Collected in real-time using Selenium-based web scraping.
3. **Tweet Data**
 - Each tweet is stored with the following details:
 - Stock Symbol (e.g., RCAT, DKNNG)
 - Tweet Text (e.g., "\$LYFT is looking strong after earnings!")
4. **Sentiment Classification**
 - The tweet text is analyzed using the VADER sentiment analysis model.
 - Each tweet is classified as Positive, Negative, or Neutral based on sentiment scores.
5. **Final Processed Data**
 - A structured CSV file containing stock symbols, tweet text, and sentiment classification for further insights and analysis.

METHODOLOGY

The process of analyzing stock-related discussions on X (formerly Twitter) involved three key steps:

1. **Web Scraping for Stock Mentions**

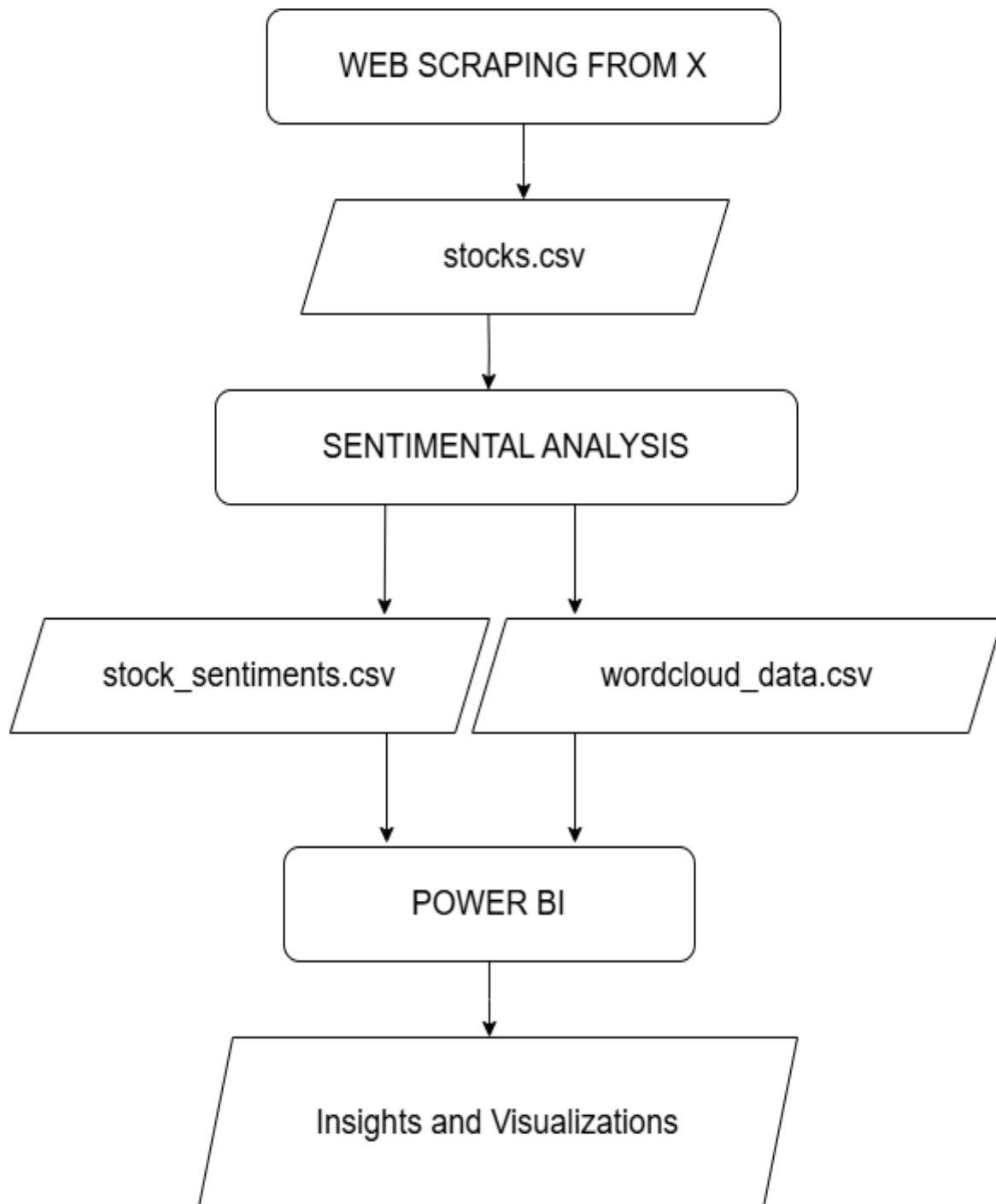
The first step was to collect real-time data from X by scraping posts mentioning specific stock tickers (e.g., \$RCAT, \$DKNNG). A web scraping script was developed using Selenium to navigate X, extract relevant tweets, and store them in a structured CSV file. Each entry in the dataset included the stock name and the corresponding post.
2. **Sentiment Analysis of Stock Mentions**

After gathering the tweets, sentiment analysis was performed using the VADER (Valence Aware Dictionary and Sentiment Reasoner) model. This step involved processing the text of each tweet to determine its sentiment—whether it was positive, negative, or neutral. The results were added to the CSV file, providing a structured dataset with sentiment labels for each stock-related post.
3. **Data Visualization and Insights in Power BI**

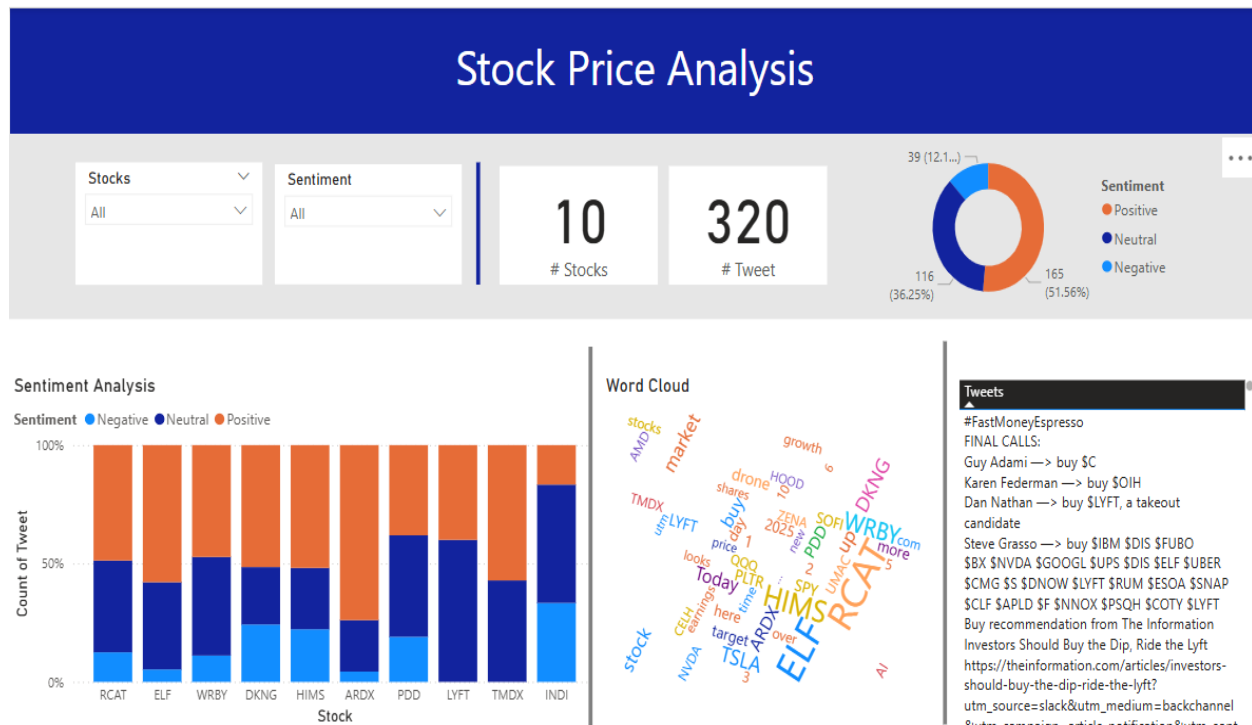
The sentiment-labeled dataset was then imported into Power BI for further analysis. Using visualization tools, trends and patterns in stock sentiment were examined, enabling deeper insights into market sentiment shifts, stock performance correlation, and potential trading signals based on public sentiment.

METHODOLOGY ALGORITHM

The following structured approach provides a comprehensive workflow for tracking stock discussions, analyzing sentiment, and deriving meaningful insights through data visualization.

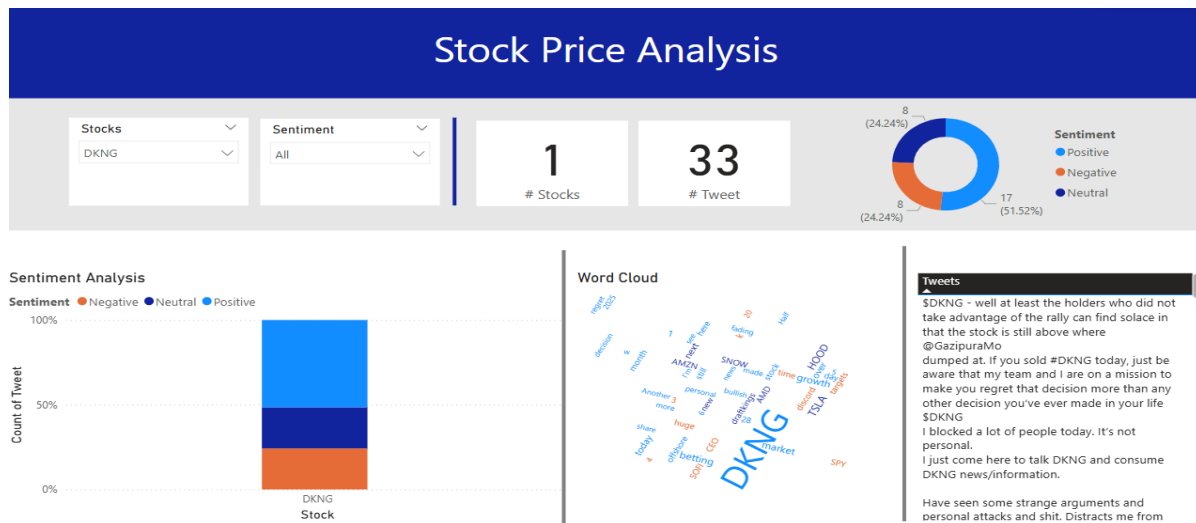


Power Bi Dashboard for analysis:



- ☐ **Total Tweets:** 320 tweets mention various stocks.
- ☐ **Stocks Analyzed:** 10 different stocks, including WRBY, RCAT, DKNG, HIMS, and TMDX.
- ☐ **Sentiment Distribution:**
 - 51.56% of tweets have a **positive** sentiment.
 - 36.25% are **neutral** tweets.
 - The remaining 12.19% have a **negative** sentiment.
- ☐ **Stock Sentiment Analysis:**
 - Stocks like **RCAT, ELF, WRBY, and DKNG** have a mix of positive and neutral sentiments.
 - **INDI and TMDX** show a higher proportion of **positive** sentiment.
 - Some stocks have notable **negative** sentiment, possibly indicating concerns.
- ☐ **Word Cloud Highlights:** Key terms discussed include **TMDX, market, stock, and investors** among others.
- ☐ **Tweet Mentions:** Discussions include **buy recommendations** for certain stocks like LYFT and IBM.

1. DKNNG STOCK (DraftKings Inc.)



- 33 tweets mention this stock.
- 52% of these tweets have a positive sentiment.
- Users are discussing key terms related to this stock, including "DKNNG," "TSLA," and "CEO."
- "CEO" is tagged to 'Positive' sentiment, further supported by a tweet where CEO Rabins states, "We are ahead of growth targets."

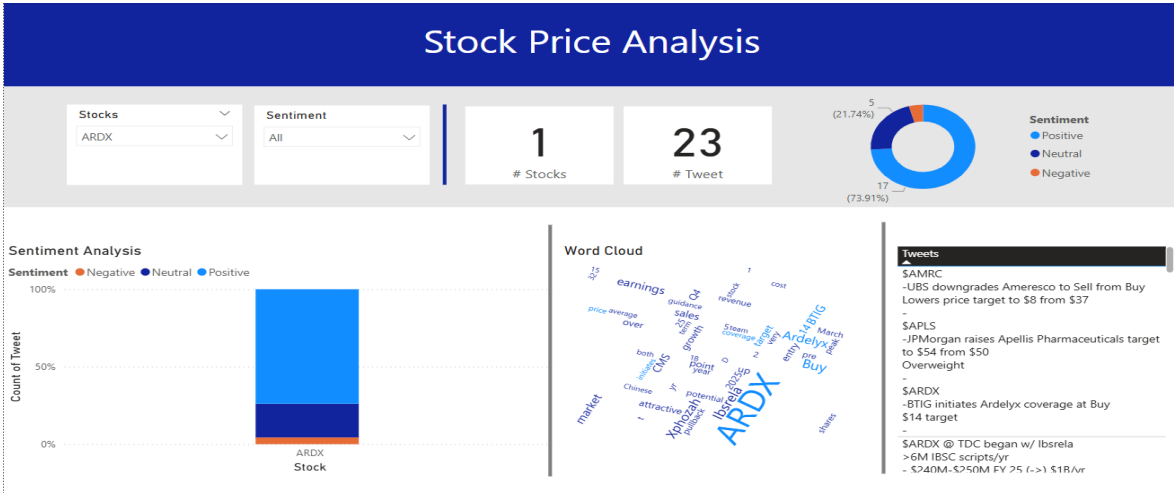
Positive Sentiment Analysis:

- One tweet mentions that even though some investors didn't sell their \$DKNG shares when the price was high, they can still feel good because the stock price remains above where another investor (@GazipuraMo) sold it. This suggests that holding onto the stock was not a bad decision compared to selling too early.
- Another tweet highlights that Benchmark maintains its "Buy" rating on \$DKNG, reinforcing confidence in the stock's future growth.
- Some tweets discuss how DraftKings' new platform, "DraftKings Predict," has been registered with the National Futures Association, hinting at potential expansion and regulatory approvals.
- Investors are optimistic as they believe DKNNG's short-term volatility does not affect its long-term potential.

Improvement Suggestion Based on Negative Tweets:

- Some negative tweets mention concerns about DKNNG's high competition and market volatility. DraftKings can address these concerns by highlighting its unique market advantages, upcoming innovations, and strategies to maintain long-term stability.
- **Overall, people are happy with DKNNG stocks, driven by positive ratings, executive confidence, and market developments.**

2. ARDX STOCK (Ardelyx Inc.)



- 23 tweets mention this stock.
- 73.91% of these tweets have a positive sentiment.
- Users are discussing key terms related to this stock, including "ARDX," "earnings," "sales," and "guidance."
- “Buy” is tagged to ‘Positive’ sentiment, as seen in a tweet where BTIG initiates coverage on Ardelyx with a “Buy” rating and a \$14 price target.

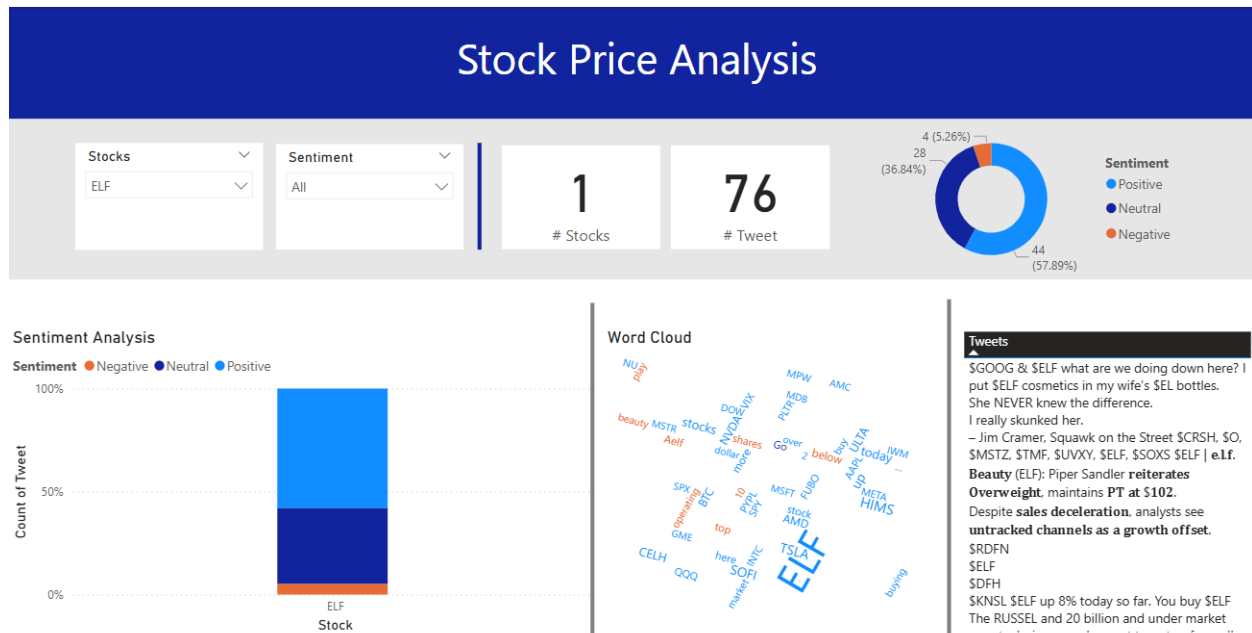
Positive Sentiment Analysis:

- One tweet mentions that BTIG has initiated coverage on \$ARDX with a “Buy” rating and a price target of \$14, showing strong confidence in the stock’s potential growth.
- Another tweet highlights that JPMorgan has increased its price target for Apellis Pharmaceuticals, indicating a positive trend in the biotech sector that could benefit \$ARDX.
- Investors are discussing Ardelyx’s partnership with IBSrela, which has reportedly exceeded 6 million prescriptions per year, reflecting strong product demand.
- Some tweets mention revenue projections for \$ARDX, with expected fiscal year earnings reaching up to \$1 billion, signaling optimism in its financial growth.

Improvement Suggestion Based on Negative Tweets:

- Some negative tweets mention concerns about stock volatility and potential regulatory risks. Ardelyx can address these by improving investor communication and providing clarity on future regulatory strategies.
- **Overall, investors are highly optimistic about \$ARDX due to strong analyst ratings, increasing sales, and promising revenue projections.**

3. ELF STOCK (E.L.F Beauty, Inc.)



- 76 tweets mention this stock.
- 57.89% of these tweets have a positive sentiment.
- Users are discussing key terms related to this stock, including "ELF," "beauty," "stocks," and "PT \$102."
- One positive tweet highlights that analysts maintain an "Overweight" rating on ELF stock and see potential growth despite a slowdown in sales.

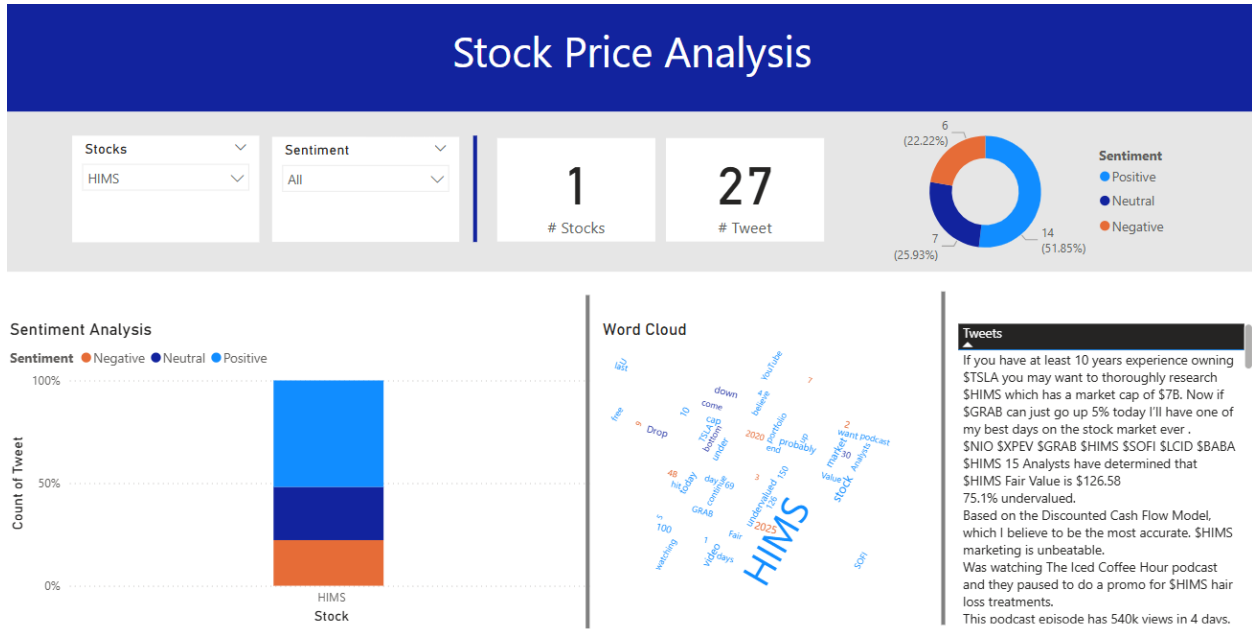
Positive Sentiment Analysis:

- A tweet mentions that Piper Sandler reiterated an Overweight rating for ELF with a price target of \$102, reinforcing investor confidence.
- Some tweets discuss ELF's ability to find growth through untracked channels, suggesting that the company is leveraging new sales or marketing strategies.
- Investors are optimistic as ELF stock continues to perform well despite industry challenges, showing resilience in the market.
- Another tweet highlights ELF's strong presence in the beauty industry, positioning it as a competitive and growing brand.

Improvement Suggestion Based on Negative Tweets:

- Some negative tweets mention concerns about sales deceleration. ELF can address this by increasing transparency in its growth strategy and emphasizing its plans for continued expansion.
- **Overall, the sentiment around ELF stock is positive, with investors optimistic about its growth potential and strategic positioning.**

4. HIMS STOCK (Hims & Hers Health, Inc.)



- 27 tweets mention this stock.
- 51.85% of these tweets have a positive sentiment.
- Users are discussing key terms related to this stock, including "HIMS," "market cap," and "fair value."
- Some tweets highlight that analysts have determined HIMS' fair value to be \$126.58, indicating potential undervaluation.

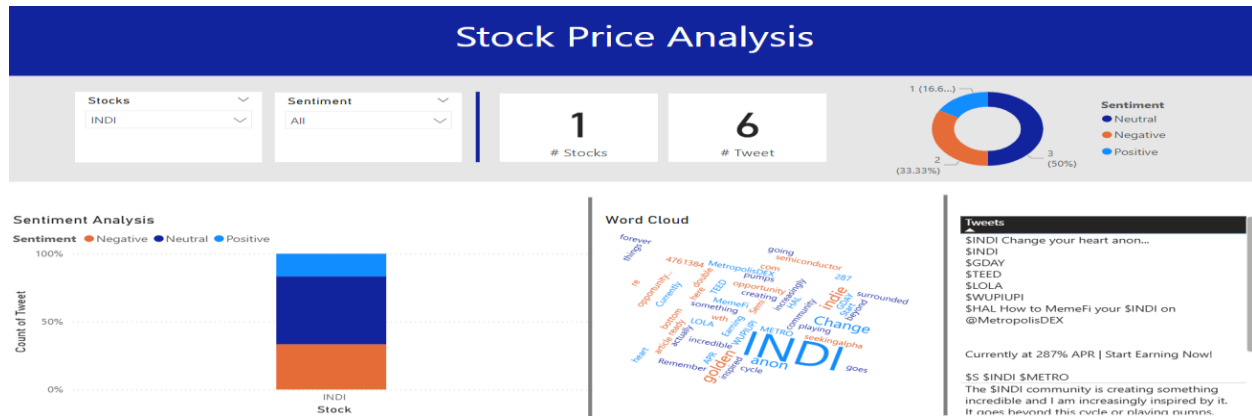
Positive Sentiment Analysis:

- One tweet discusses how HIMS' marketing is highly effective, with promotions appearing on popular podcasts like The Iced Coffee Hour, which had 540K views in four days.
- Investors see potential in HIMS due to its strong market presence and advertising strategy, which could lead to higher future valuation.
- The mention of a \$7B market cap and a possible fair value of \$126.58 has sparked optimism among investors.
- Some tweets highlight that HIMS is seen as an undervalued stock, creating a buying opportunity.

Improvement Suggestion Based on Negative Tweets:

- Some negative tweets mention **concerns about competition and market uncertainty**. HIMS can address this by emphasizing its unique marketing strategies, customer retention plans, and potential business expansions to strengthen investor confidence.
- **Overall, investors are optimistic about HIMS due to its strong marketing, analyst valuations, and potential for future growth.**

5. INDI STOCK (indie Semiconductor, Inc.)



- 6 tweets mention this stock.
- 33% of these tweets have a positive sentiment, 50% are neutral, and 16% are negative.
- Users are discussing key terms related to this stock, including "INDI," "anon," and "golden."
- "golden" is tagged to 'Positive' sentiment, supported by tweets referring to INDI as a golden opportunity.

Positive Sentiment Analysis

- One tweet suggests that the INDI community is creating something incredible, expressing strong enthusiasm for the stock's future.
- Another tweet mentions INDI as an opportunity, hinting at potential gains and investment interest.

As this stock has more negative tweets, let's do a negative sentiment analysis too.

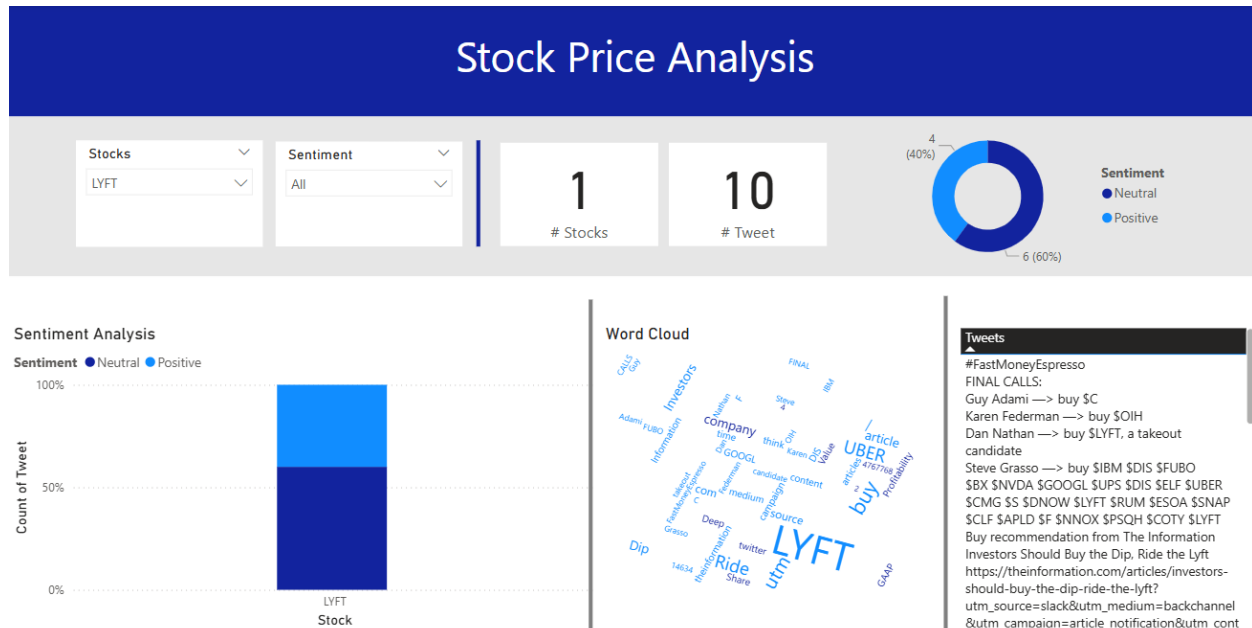
Negative Sentiment Analysis

- Some tweets express concerns about INDI's volatility and uncertain market position.
- One tweet hints at INDI being overhyped, suggesting that the stock's momentum might not last.
- Another negative tweet mentions uncertainty in the semiconductor market, which could impact INDI's performance.

Improvement Suggestion Based on Negative Tweets

- Investors are concerned about INDI's long-term stability and whether the hype is justified. The company can address these concerns by clarifying growth strategies, ensuring transparency in its future plans, and showcasing strong fundamentals.
- **Overall, discussions about INDI stock are mixed, with optimism from its community but concerns about volatility and sustainability.**

6. LYFT STOCK (Lyft, Inc.)



- 10 tweets mention this stock.
- 60% of these tweets have a positive sentiment.
- Users are discussing key terms related to this stock, including "LYFT," "buy," "ride," and "investors."
- Some tweets highlight that analysts view LYFT as a potential **takeout candidate**, indicating investor interest.

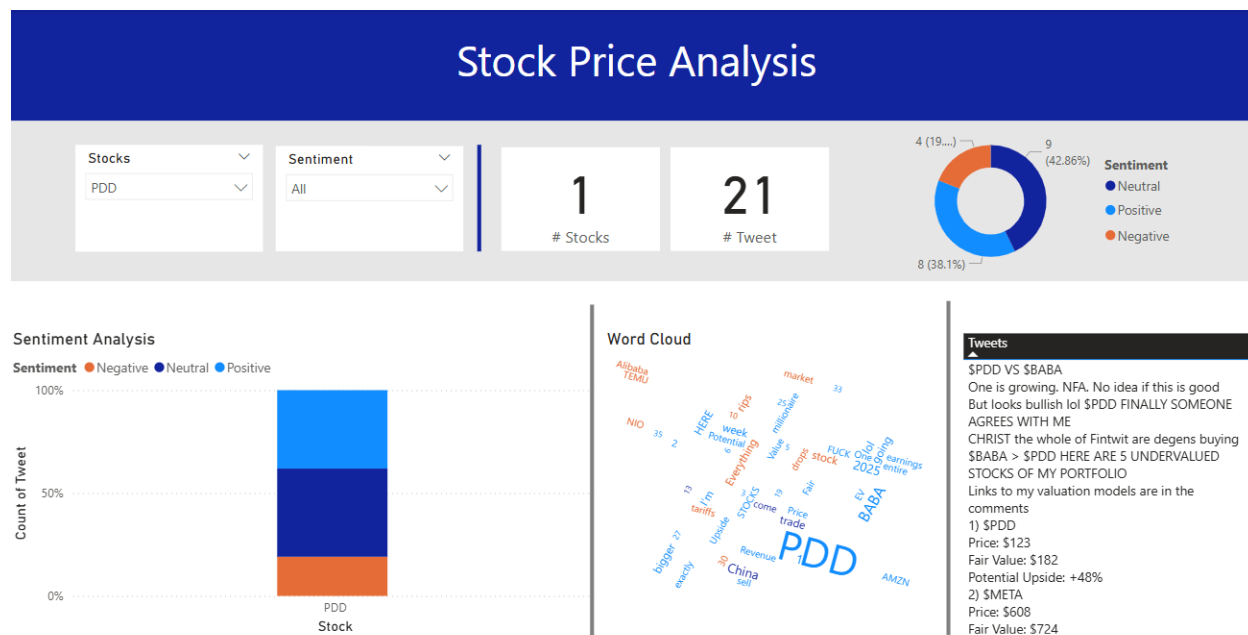
Positive Sentiment Analysis:

- Several tweets suggest buying LYFT stock, with analysts considering it an opportunity for investment.
- Tweets mention that LYFT is being recommended by financial experts as a **buy-the-dip** stock, highlighting its recovery potential.
- Some investors see LYFT's ride-sharing business model as an essential player in the industry, creating long-term value.

Improvement Suggestion Based on Neutral Tweets:

- Some neutral tweets mention industry competition and regulatory challenges. LYFT can enhance investor confidence by emphasizing strategic partnerships, cost-efficiency measures, and expansion plans.
- **Overall, investors see LYFT as a promising stock, with analysts recommending it as a buy opportunity and highlighting its potential acquisition interest.**

7. PDD STOCK (Pinduoduo Inc.)



- 21 tweets mention this stock.
- 42.86% of these tweets have a positive sentiment.
- Users are discussing key terms related to this stock, including "PDD," "BABA," "China," and "Fair Value."
- Some tweets highlight that PDD is being compared with BABA, and discussions suggest that PDD is considered an undervalued stock with strong growth potential.

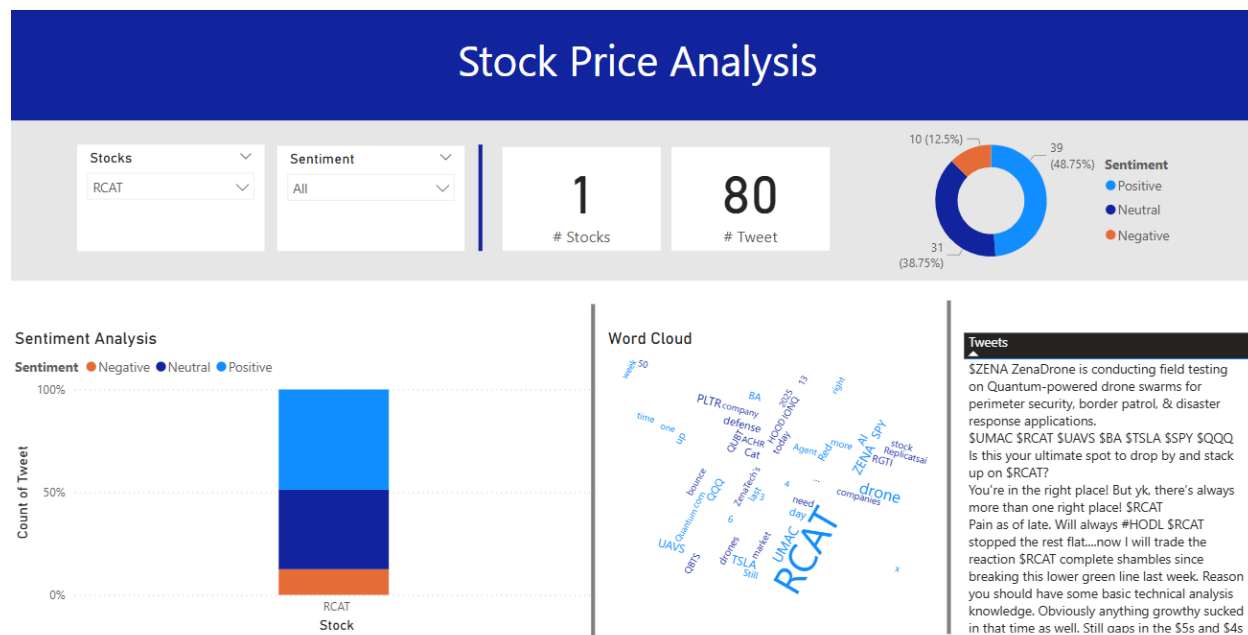
Positive Sentiment Analysis:

- Some tweets mention that PDD has a **fair value of \$182** and a **potential upside of 48%**, which attracts investors looking for undervalued stocks.
- Investors are debating whether PDD has stronger growth potential compared to BABA, with some preferring PDD for long-term gains.
- There is optimism around PDD's market expansion and revenue potential, particularly in comparison to other Chinese tech stocks.

Improvement Suggestion Based on Negative Tweets:

- Some negative tweets mention **concerns about market risks and tariffs impacting PDD's future performance**. The company can address these concerns by highlighting strategies to navigate global trade regulations and sustain its revenue growth.
- **Overall, investors are divided on PDD, but many see it as an undervalued stock with strong growth potential compared to BABA.**

8. RCAT STOCK (Red Cat Holdings, Inc.)



- 80 tweets mention this stock.
- 48.75% of these tweets have a positive sentiment.
- Users are discussing key terms related to this stock, including "RCAT," "drone," "UAVS," and "ZenaDrone."
- Many tweets highlight that ZenaDrone is conducting field tests on quantum-powered drone swarms for security and disaster response applications.

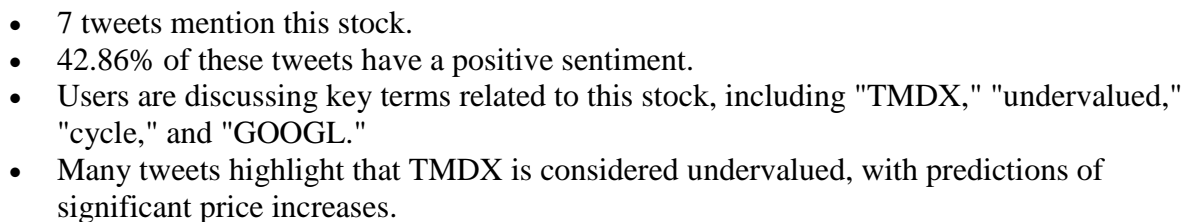
Positive Sentiment Analysis

- Some tweets mention growing optimism around RCAT's drone technology and its potential in defense and security sectors.
- Investors are discussing RCAT's increasing relevance in advanced drone solutions, boosting interest in the stock.
- The mention of potential breakout levels and technical patterns has led to positive speculation.

Improvement Suggestion Based on Negative Tweets

- Some negative tweets highlight concerns about recent price fluctuations and technical resistance levels. Investors note a need for stronger fundamental backing to sustain growth.
- RCAT can address this by showcasing more partnerships, technology advancements, and long-term growth strategies to build investor confidence.
- **Overall, investors see promise in RCAT due to its involvement in innovative drone technology, but market uncertainty and technical resistance remain concerns.**

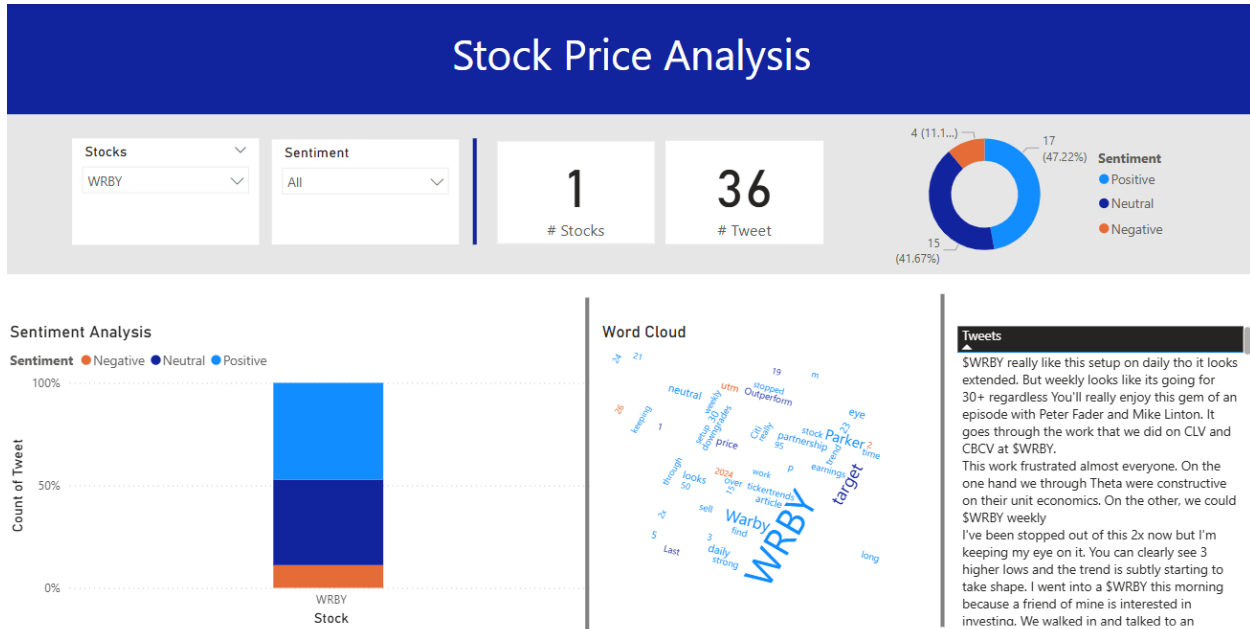
Stock Price Analysis



- Some tweets express confidence in TMDX's growth potential, with predictions of its price rising to \$130.
- Investors compare TMDX with other major stocks like GOOGL, believing it has room for substantial appreciation.
- Technical analysis discussions suggest TMDX is approaching key resistance levels, fueling positive speculation.

- Some concerns revolve around market uncertainty and the need for confirmation of a breakout above key price levels.
- TMDX can address this by reinforcing investor confidence through strong earnings reports and strategic growth initiatives.

10. WRBY STOCK (Warby Parker Inc.)



- 36 tweets mention this stock.
- 44.44% of these tweets have a positive sentiment.
- Users are discussing key terms related to this stock, including "target," "warby," and others.
- Many tweets highlight that Warby is considered undervalued, with predictions of significant price increases.

Positive Sentiment Analysis

- Some tweets express confidence in Warby's growth potential, with predictions of its price rising significantly.
- Investors compare Warby with other major stocks, believing it has room for substantial appreciation.
- Technical analysis discussions suggest Warby is approaching key resistance levels, fueling positive speculation.

Improvement Suggestion Based on Negative Tweets

- Some concerns revolve around market uncertainty and the need for confirmation of a breakout above key price levels.
- Warby can address this by reinforcing investor confidence through strong earnings reports and strategic growth initiatives.

Overall, investors have a cautiously optimistic outlook on Warby, recognizing its potential but awaiting further confirmation of sustained growth.

Future Enhancements:

1. **Automated Real-Time Data Refresh**
 - Integrate Power BI with Python queries to enable automatic data refresh every hour.
 - Provides real-time stock analysis for users and companies, improving decision-making.
2. **Premium API for Expanded Data Access**
 - Utilize the premium version of X (formerly Twitter) API instead of web scraping.
 - Ensures faster data retrieval and access to a larger dataset for more accurate sentiment analysis.
3. **Advanced Sentiment Analysis with OpenAI**
 - Replace VADER with LLM models for a more nuanced sentiment analysis.
 - Improves accuracy in detecting complex financial sentiments and stock-related discussions.

Code and datasets for above project:

1. web_scraping_code.py
2. stocks_tweets.csv
3. sentimental_analysis.py
4. stock_sentiment.csv
5. wordcloud_data.py
6. wordcloud_data.csv
7. Power_BI.pbi