Data_Analysis_on_wallstreetbets_using_redditAPI

Introduction:

This report outlines the progress made in analyzing sentiments expressed in discussions on the "wallstreetbets" subreddit. The project involved scraping data from Reddit, performing sentiment analysis, and visualizing the results to uncover insights that could inform trading strategies.

Data Collection and Preprocessing:

The data was collected using the Reddit API, focusing on key attributes such as:

- ➔ **Title:** The title of each post.
- ➔ **Selftext:** The content of the post.
- ➔ **Upvote Ratio:** The ratio of upvotes to total votes.
- ➔ **Ups:** Total number of upvotes.
- ➔ **Downs:** Total number of downvotes.
- ➔ **Sentiment Score:** A numerical representation of sentiment derived from text analysis.
- ➔ **Sentiment Label:** Classification of sentiment (positive, negative, neutral).
- ➔ **Stock Mentions:** Stocks mentioned within the post.

The dataset was cleaned to remove null values and irrelevant columns, resulting in a structured CSV file for analysis.

Code Overview:

The project consists of two main Jupyter notebooks:

Notebook 1: Data Scraping

- ➔ **Libraries Used:** `requests`, `pandas`, `json`
- ➔ **Functionality:** Connects to the Reddit API to scrape posts from "wallstreetbets" and stores relevant data in a structured format.

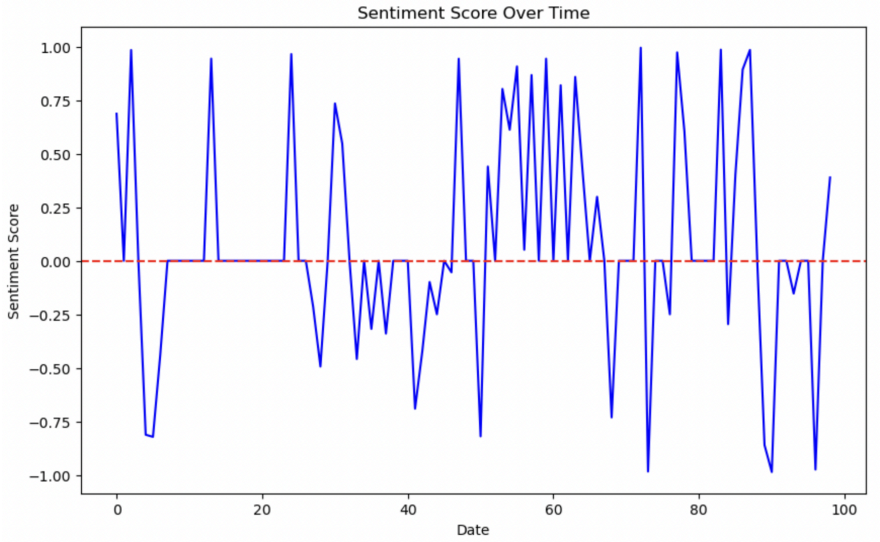
Notebook 2: Data Analysis and Visualization

- ➔ Libraries Used: `pandas`, `numpy`, `matplotlib`, `seaborn`, `nltk`
- ➔ Functionality:
  - ◆ Cleans and preprocesses the scraped data.
  - ◆ Conducts sentiment analysis using NLP techniques to classify sentiments.
  - ◆ Generates visualizations to illustrate key findings.

Visualizations:

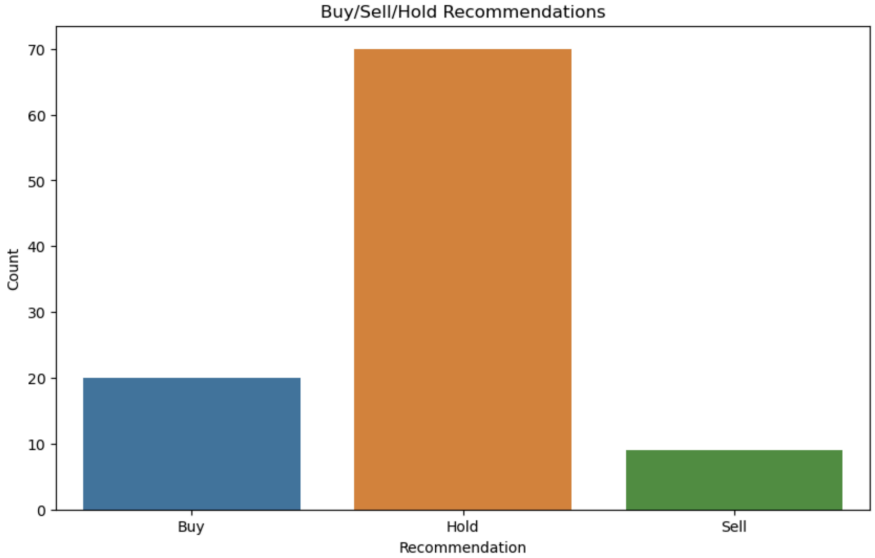
1. Sentiment Score Over Time

This graph shows the sentiment score over time. A positive score indicates a generally optimistic sentiment, while a negative score indicates pessimism. The red dashed line represents neutral sentiment.



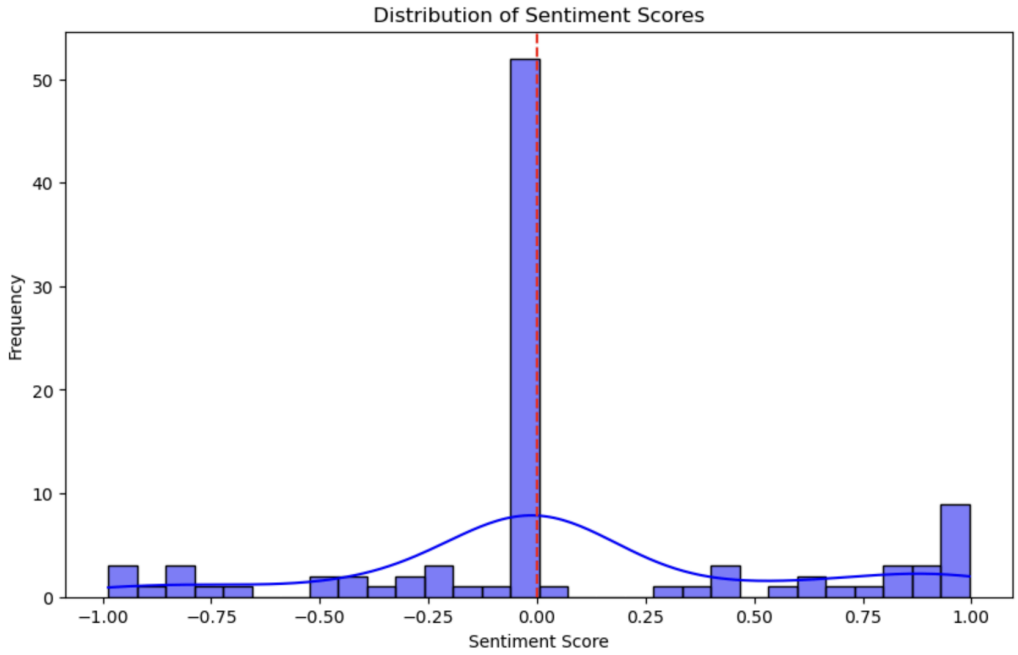
2. Buy/Sell/Hold Recommendations

This graph displays the distribution of trading signals based on sentiment scores, categorizing posts into buy, sell, and hold recommendations.



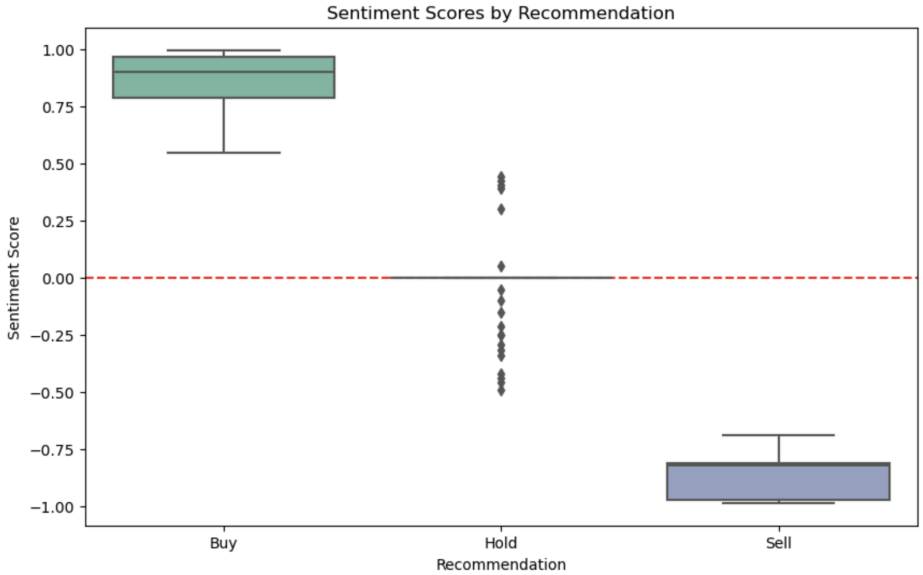
3. Distribution of Sentiment Scores

This histogram illustrates the distribution of sentiment scores across all analyzed discussions, with a density plot (KDE) showing concentration and a neutral sentiment line in red.



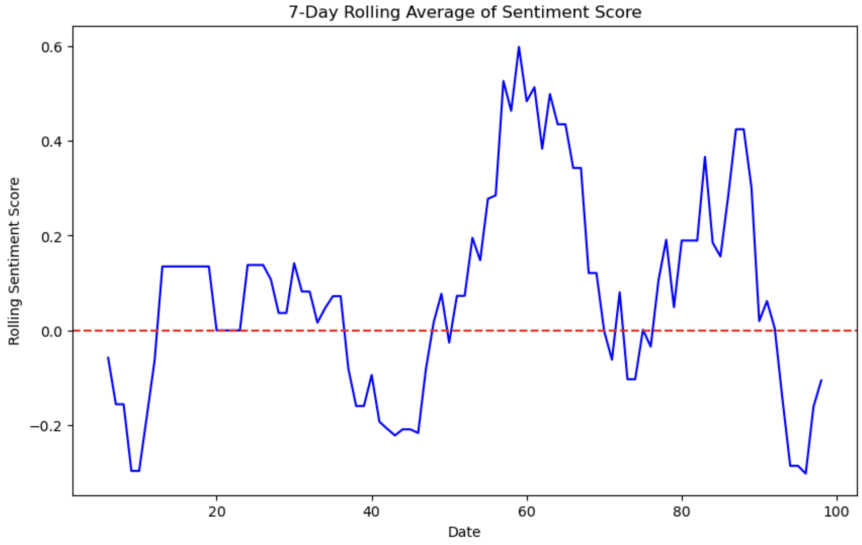
4. Sentiment Scores by Recommendation

The box plot depicts the distribution of sentiment scores for each recommendation (Buy, Sell, Hold), allowing us to visualize how sentiment varies with different trading signals.



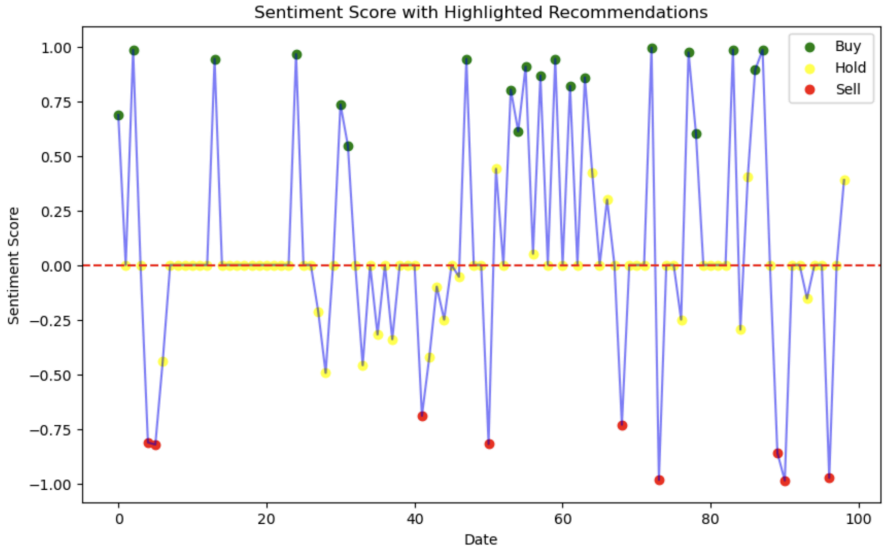
5. 7-Day Rolling Average of Sentiment Score

This graph presents the 7-day rolling average of sentiment scores, smoothing out fluctuations to reveal underlying trends.



6. Sentiment Score with Highlighted Recommendations

This graph shows sentiment scores with highlighted trading recommendations, enabling us to see how sentiment correlates with Buy/Sell/Hold signals.



Key Findings:

- ➔ Average Sentiment Score: 0.08
- ➔ Median Sentiment Score: 0.00

The analysis indicates that:

- ➔ Positive sentiments correlate with buy signals, suggesting that optimistic discussions may precede favorable market movements.
- ➔ Negative sentiments often align with sell recommendations, indicating caution among traders during periods of pessimism.
- ➔ A significant portion of posts exhibited neutral sentiments, reflecting uncertainty within the community regarding specific stocks or market conditions.

Conclusion:

The findings from this analysis provide valuable insights into potential trading signals based on community sentiments expressed on "wallstreetbets." The visualizations enhance our understanding of these sentiments and their potential impact on trading behaviors.

Future Work:

Future analyses could explore:

- ➔ Expanding the dataset over longer periods for deeper insights.
- ➔ Implementing advanced NLP techniques for improved sentiment classification.
- ➔ Correlating sentiment data with real-time stock price movements for actionable trading strategies.

This report encapsulates our progress in analyzing Reddit sentiments related to stock trading and provides a foundation for further exploration in this dynamic field.