

Subreddit Language Classification

Google
r/WallStreetBets & r/Stocks

For Sales, Service, and Support

Contents

Topics Covered

1

Business Problem

2

Background

3

Data Exploration

4

Insights & Recommendations

Business Problem

- High-profile client in the debt relief space
- Tasked by Google's Sales, Service, and Support team to target a demographic of customers in need of debt relief
- Solution: Utilize Natural Language Processing (NLP) with the Google Home, Google search, and other integrations to target users of r/WallStreetBets



Potential Market Size:
1.2M converted customers (10%)

Background



r/WallStreetBets

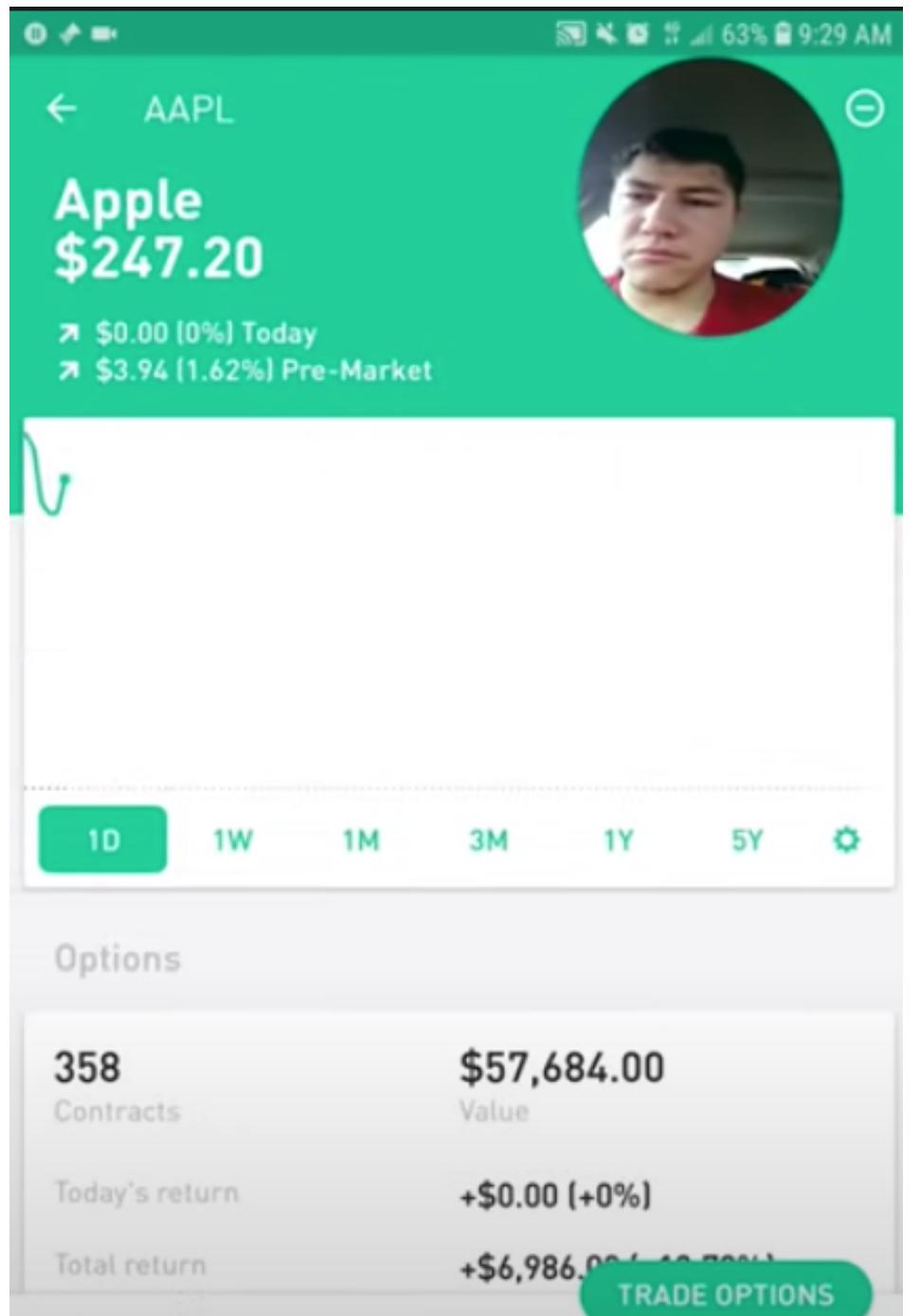
- Traders with high risk tolerance
- Users irresponsibly utilize options and margin for leverage
- Subreddit contains its own unique language



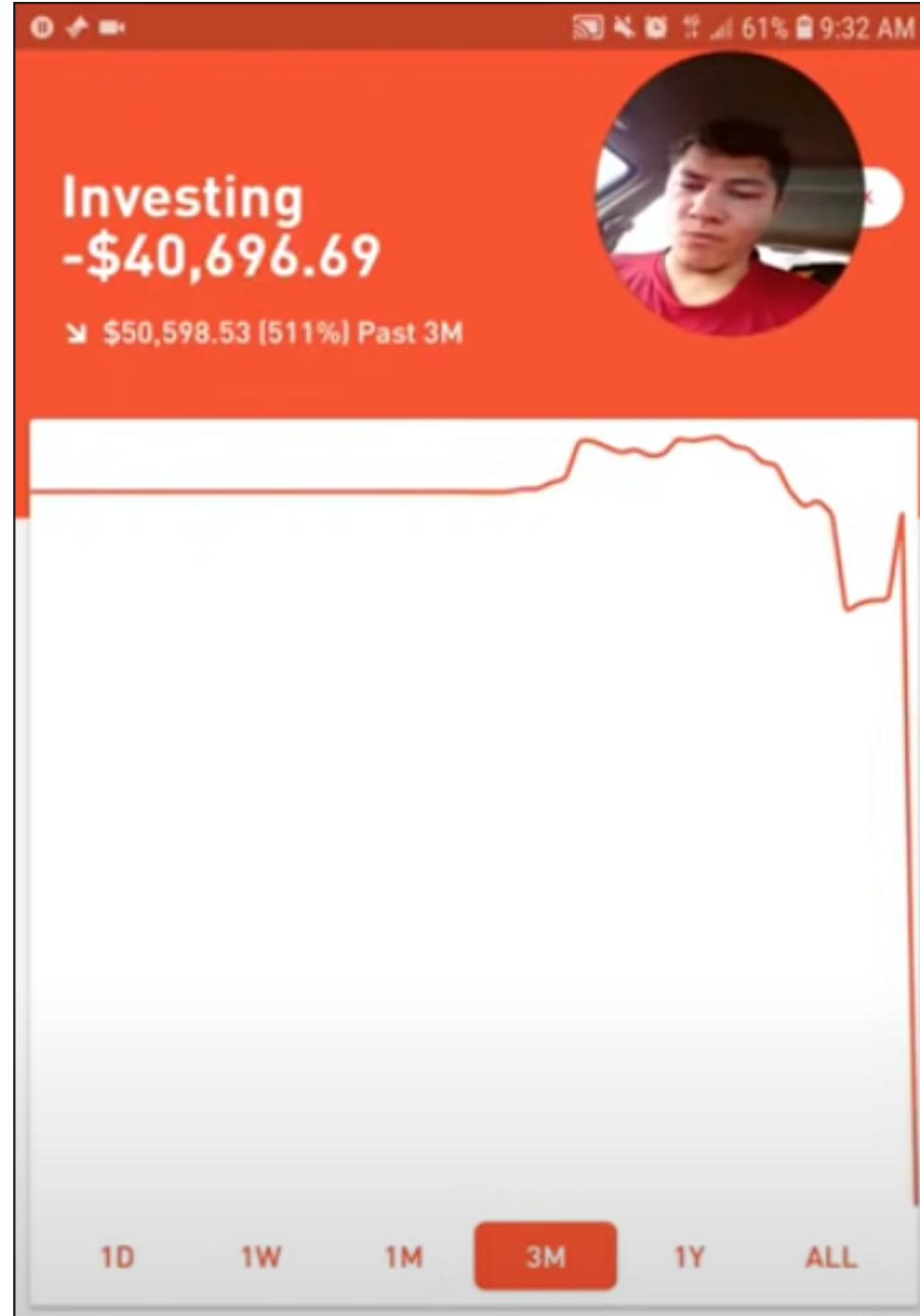
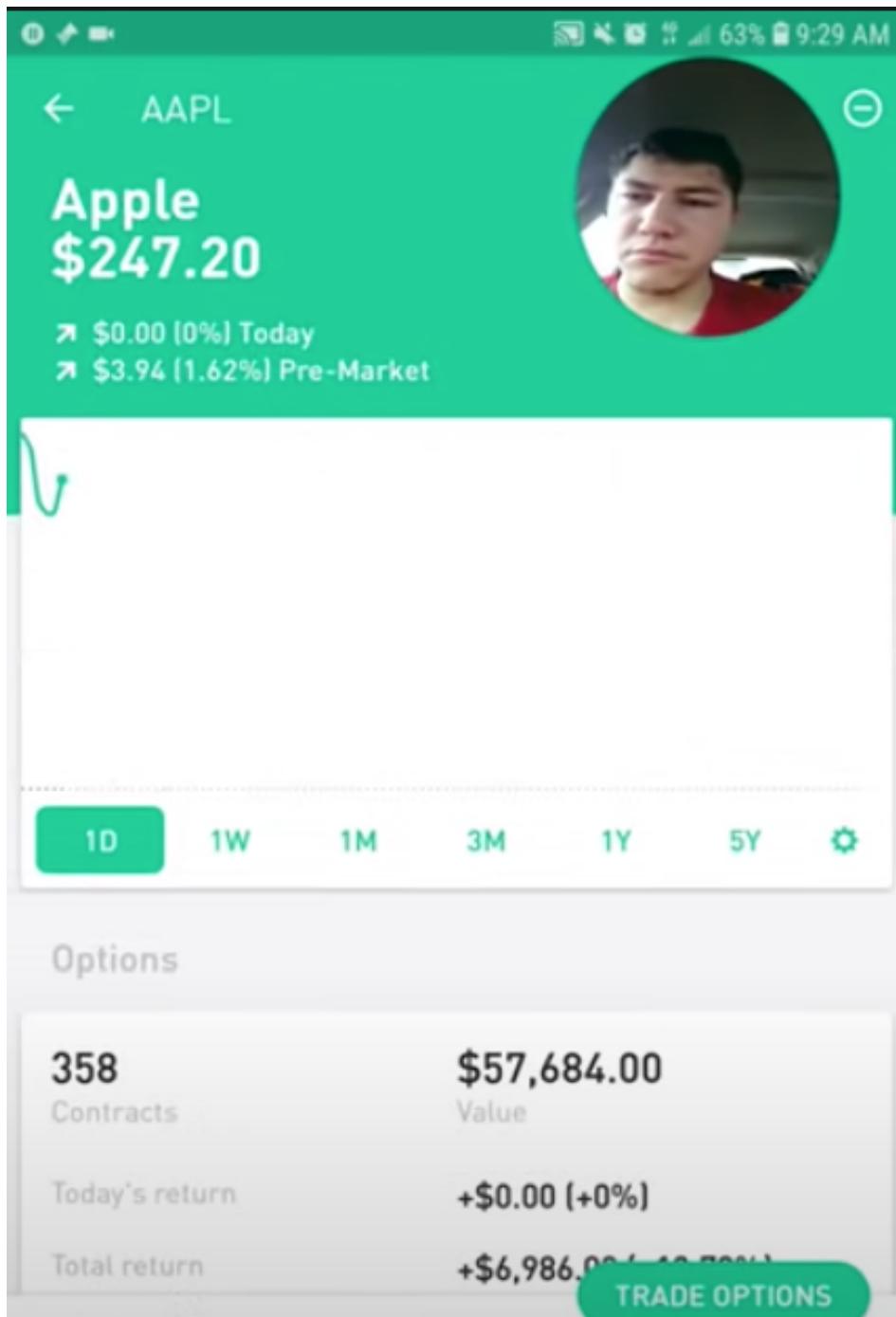
r/Stocks

- Individual investors
- Novice investors looking for assistance
- Subreddit contains basic investment jargon

Guh



Guh



Background



r/WallStreetBets

- Traders with high risk tolerance
- Users irresponsibly utilize options and margin for leverage
- Subreddit contains its own unique language



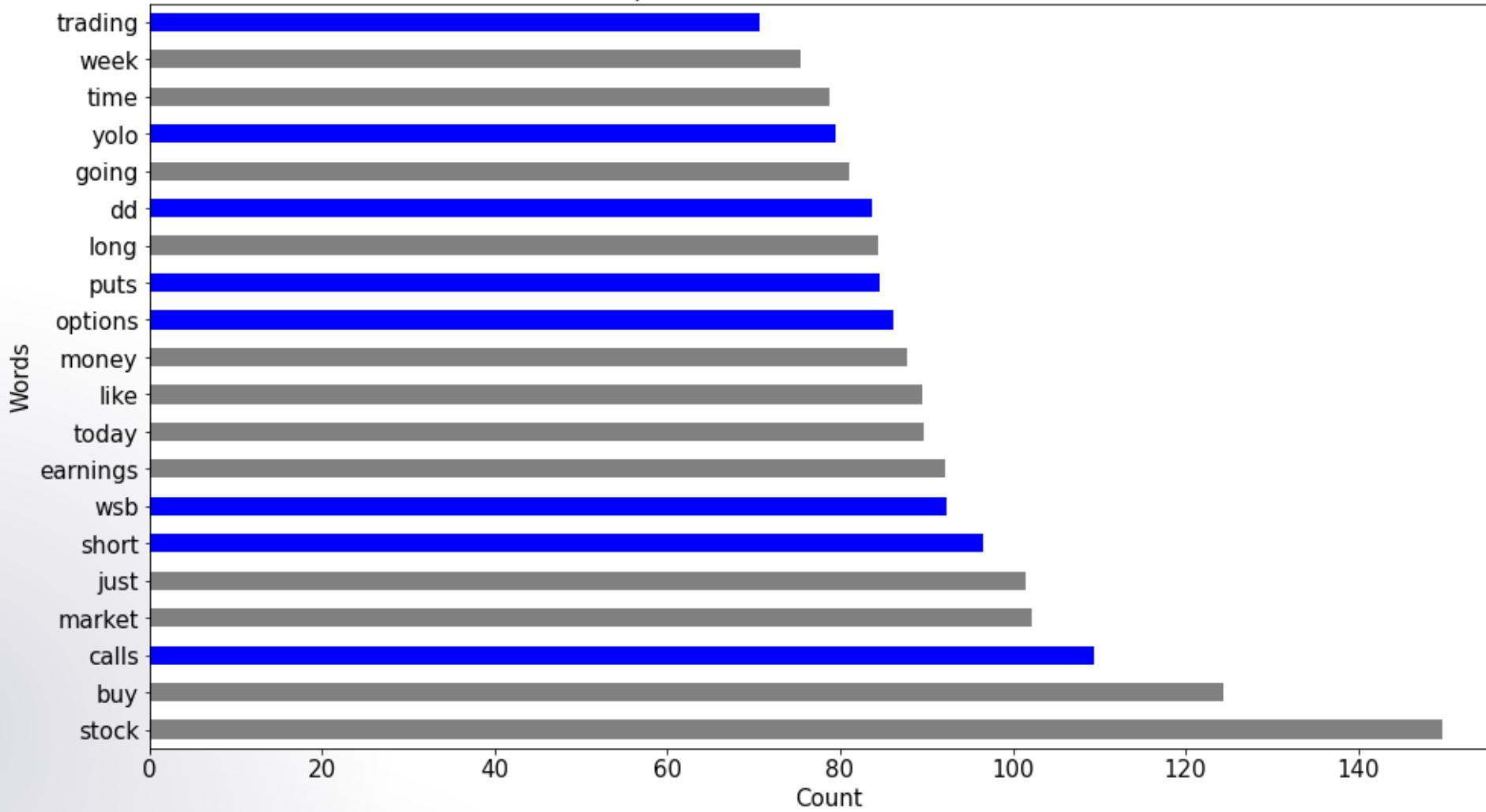
r/Stocks

- Individual investors
- Novice investors looking for assistance
- Subreddit contains basic investment jargon

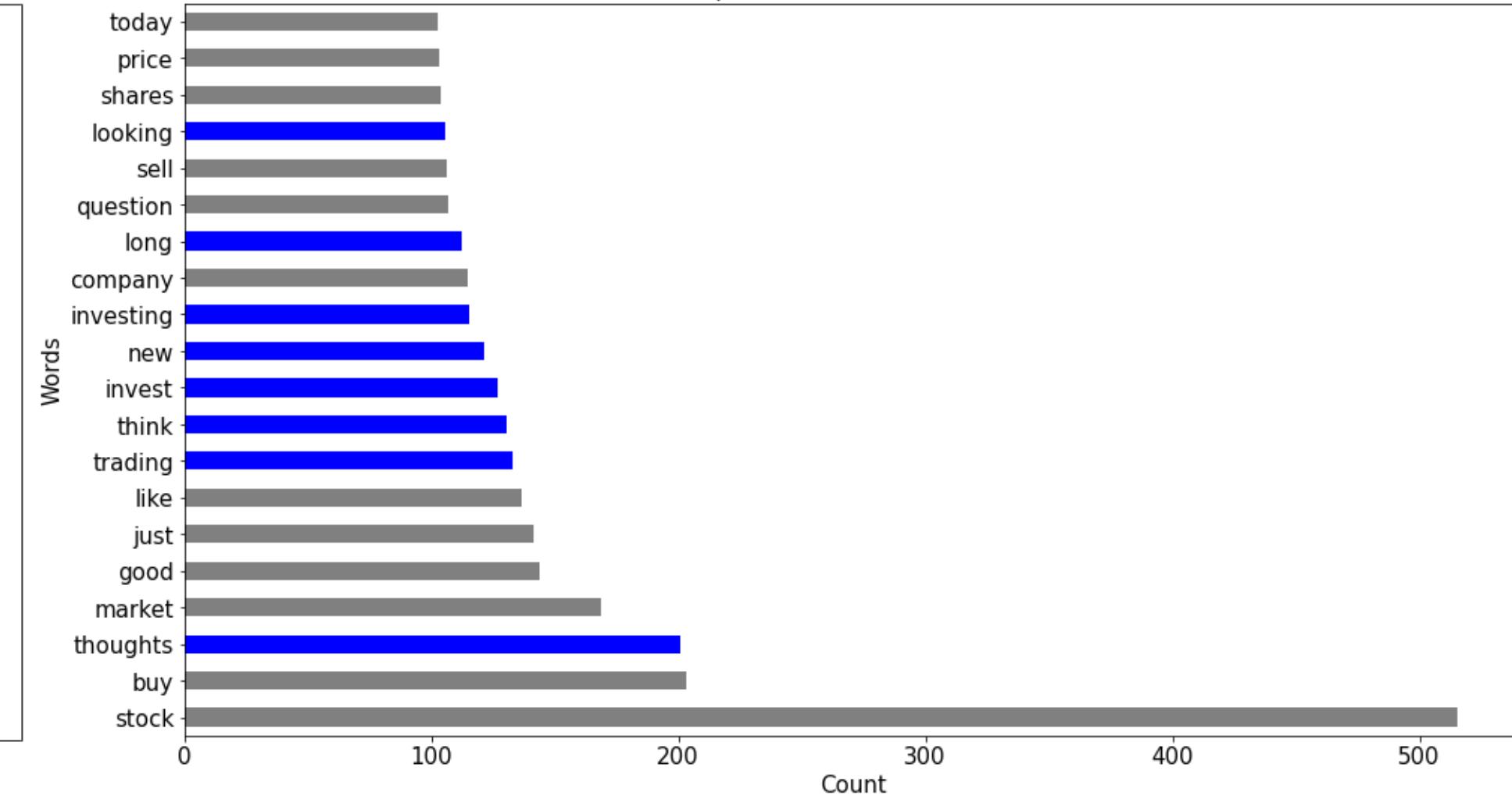
Common Words



Top r/WallStreetBets Words Used



Top r/Stocks Words Used



Sub(tle)reddit Differences

r/WallStreetBets



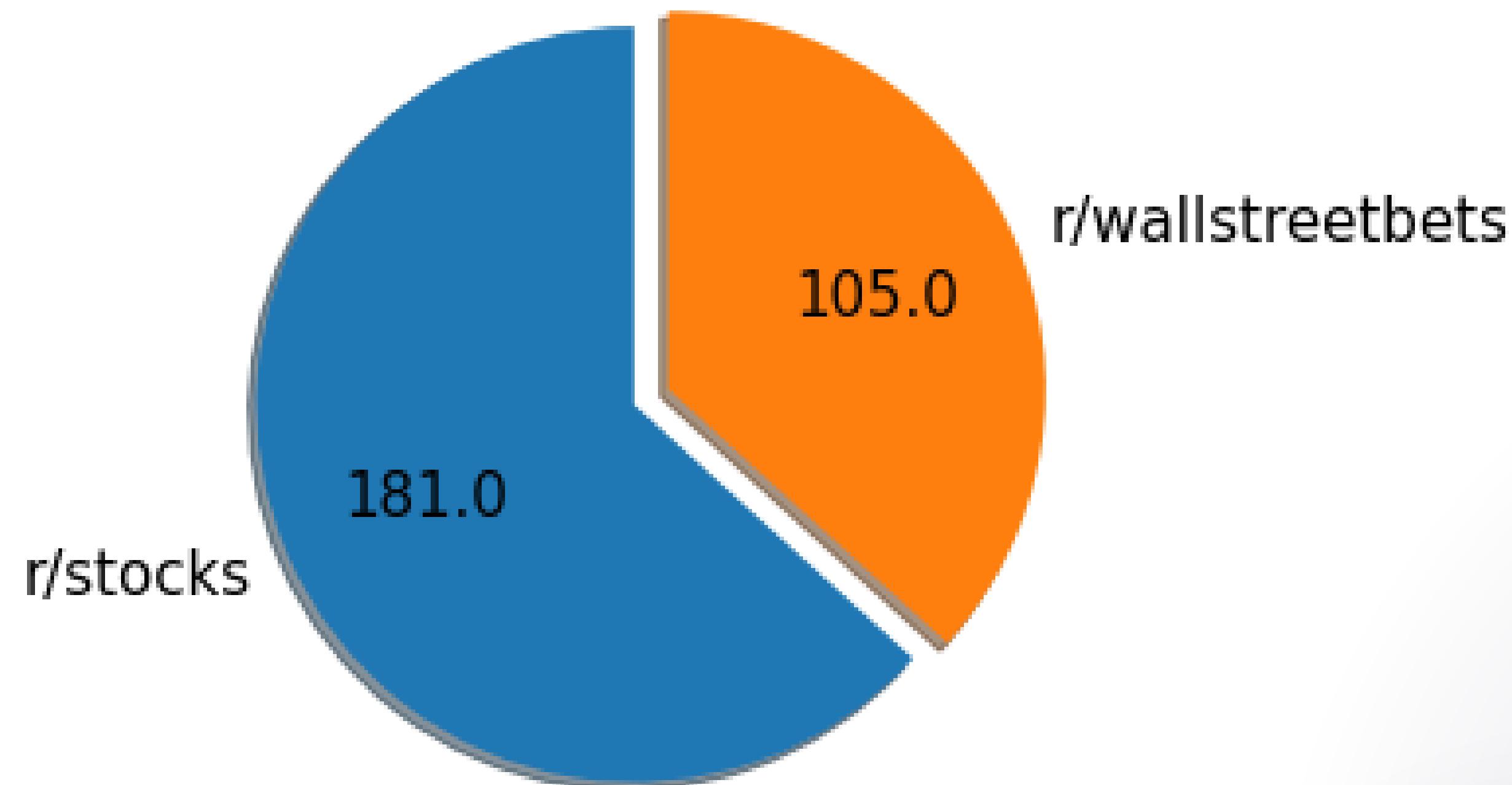
- Rampant profanity and derogatory language
 - "Calls", "Puts", and other language that alludes to a high risk tolerance
 - Other language nuances ("tendies", 'YOLO')
-

r/Stocks



- Language that suggests that users take a long-term investing strategy
- Many questions asking for strategy assistance, likely from novice investors
- 26% longer average post length

Average Text Length by Subreddit

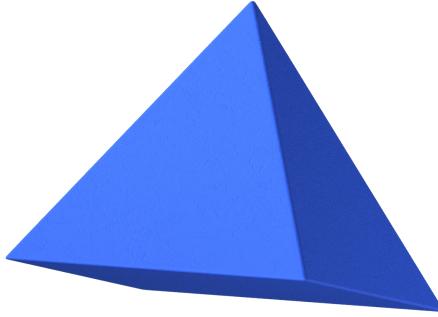


Insights

Clean / non-profanity words that were identified by the XGBoost model as important words in determining the classification between r/wallstreetbets and r/stocks, and are evidently related to the r/wallstreetbets community, include:

- dd
- moon
- meme
- yolo
- tendies
- wsb

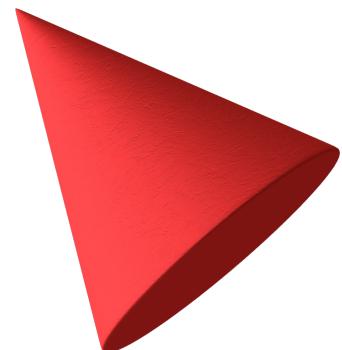
Recommendations & Next Steps



Implementation

Utilize model-identified keywords to target wallstreetbettters on Google Home and other platforms

- dd
- moon
- meme
- yolo
- tendies
- wsb



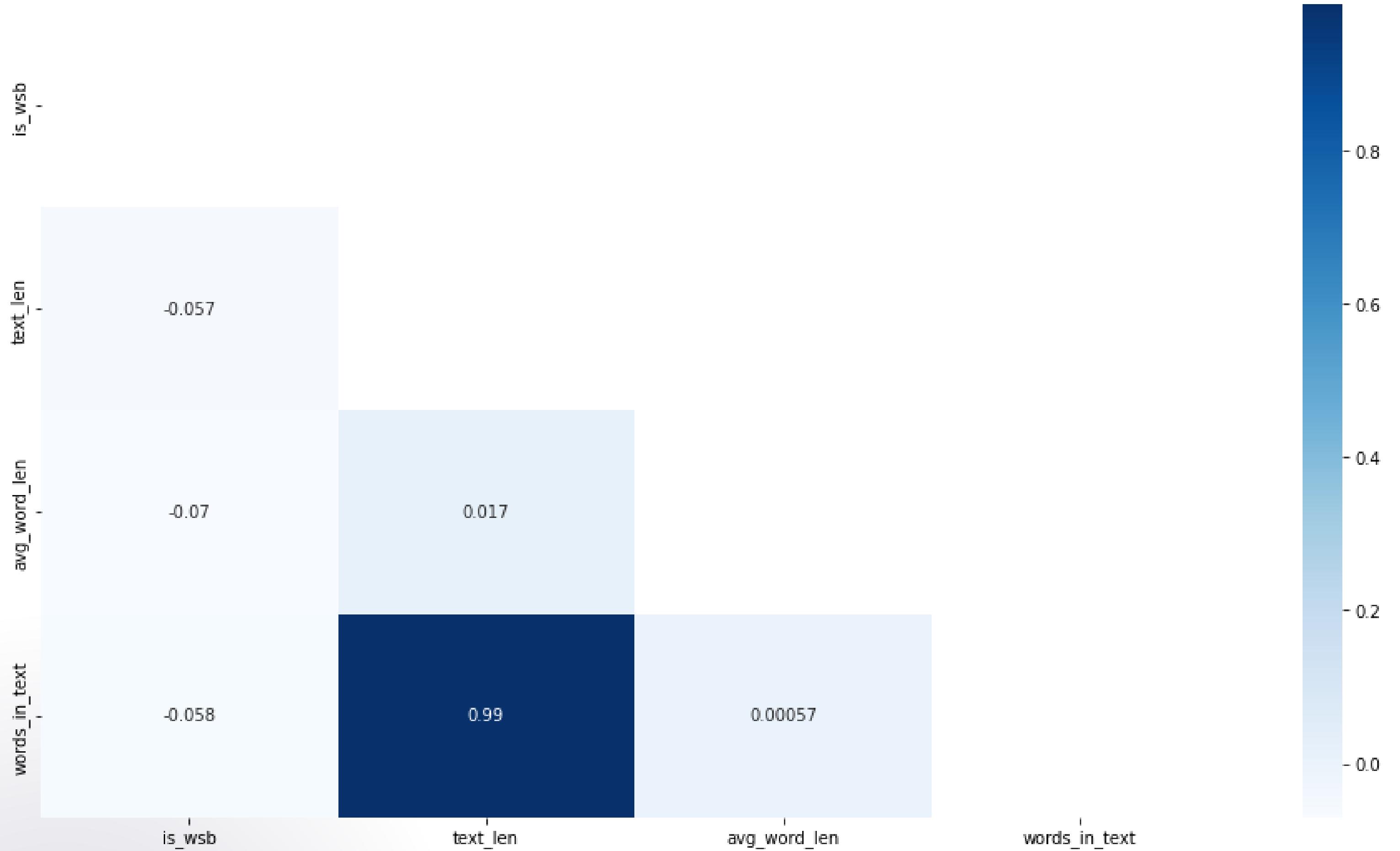
Contextual Considerations

"Moon", for example, can only work for this problem in the context of finance

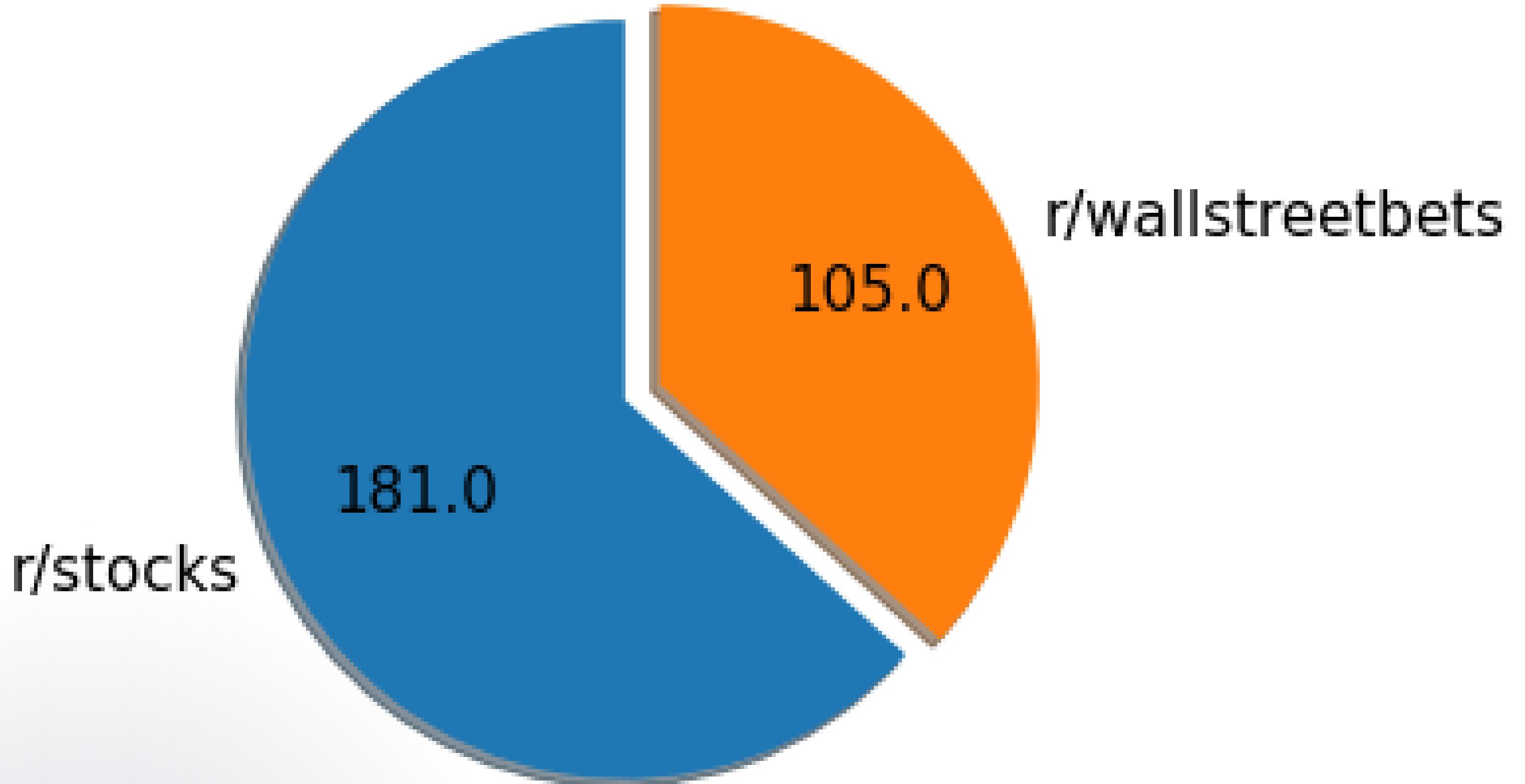
Thank You

Appendix

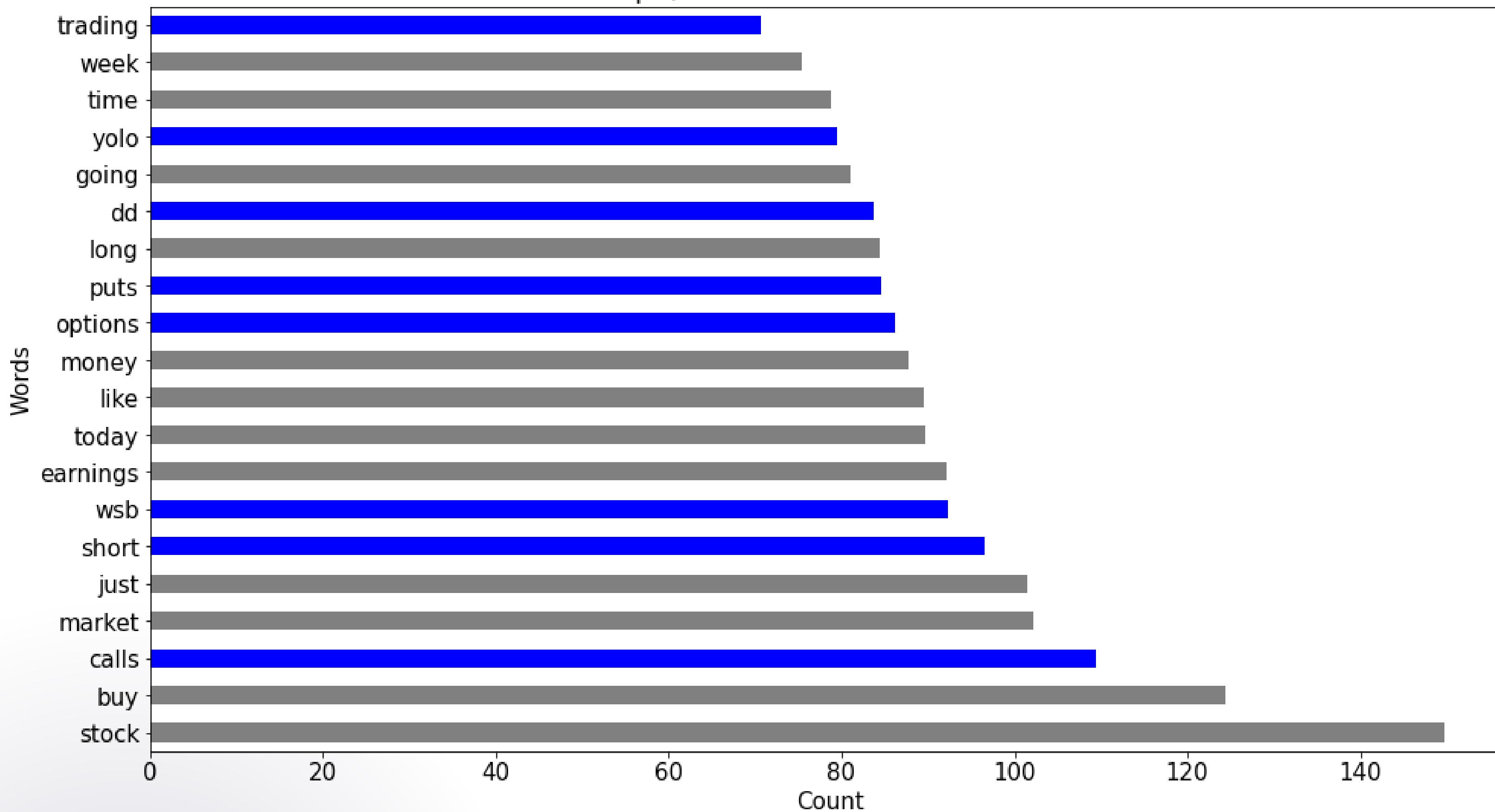
Correlation Heatmap of Subreddit Text Features



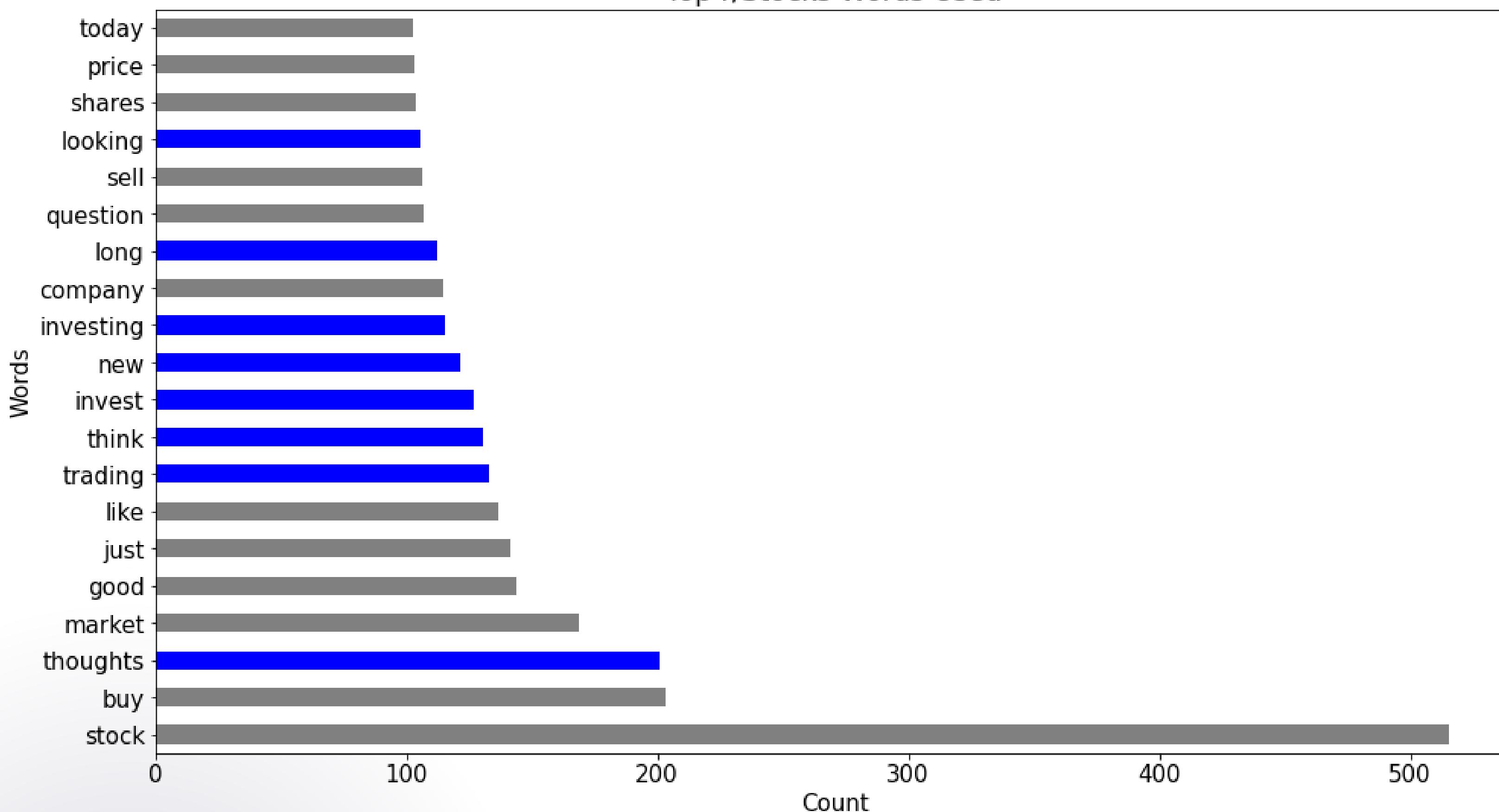
Average Text Length by Subreddit



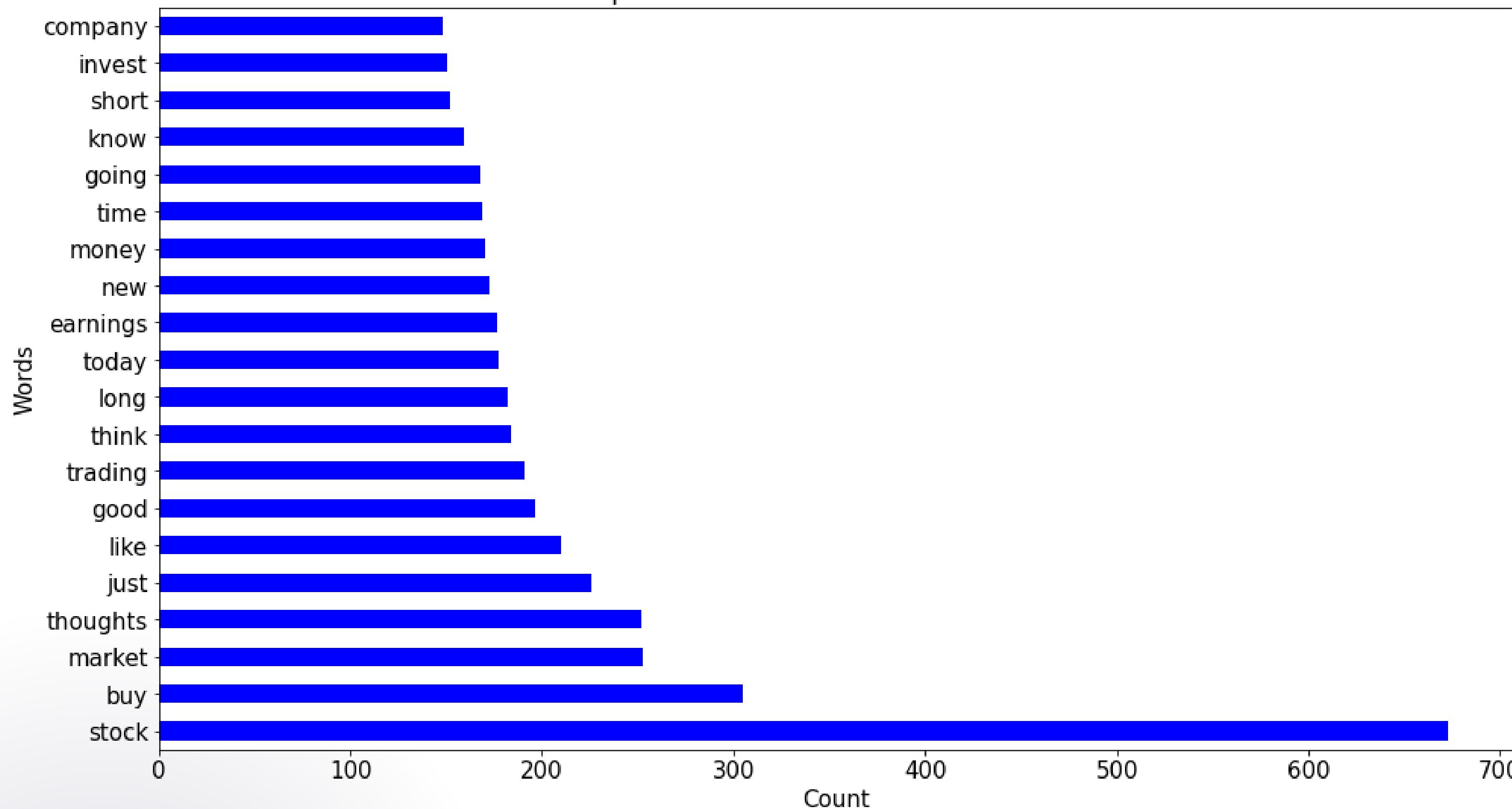
Top r/WallStreetBets Words Used



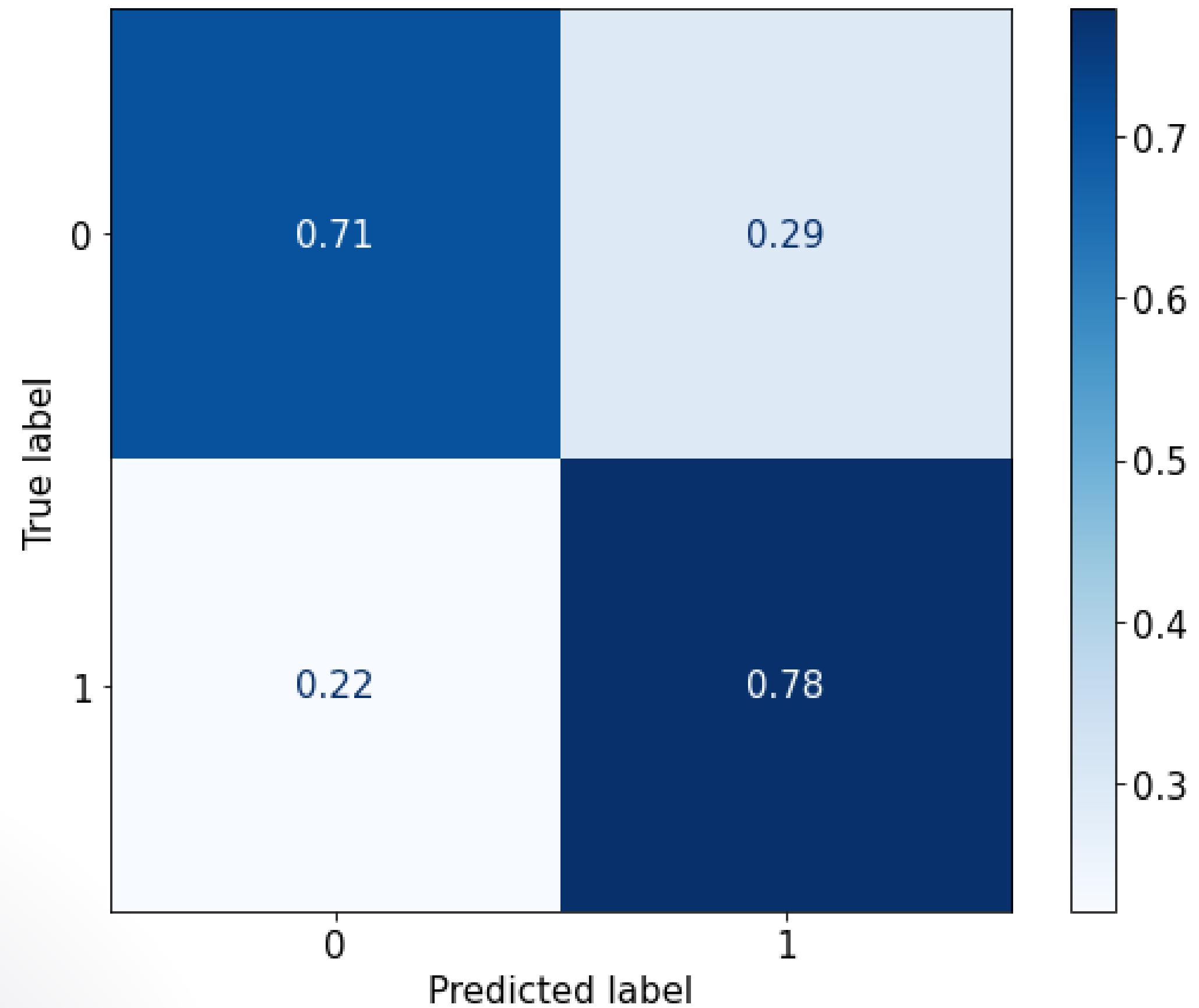
Top r/Stocks Words Used



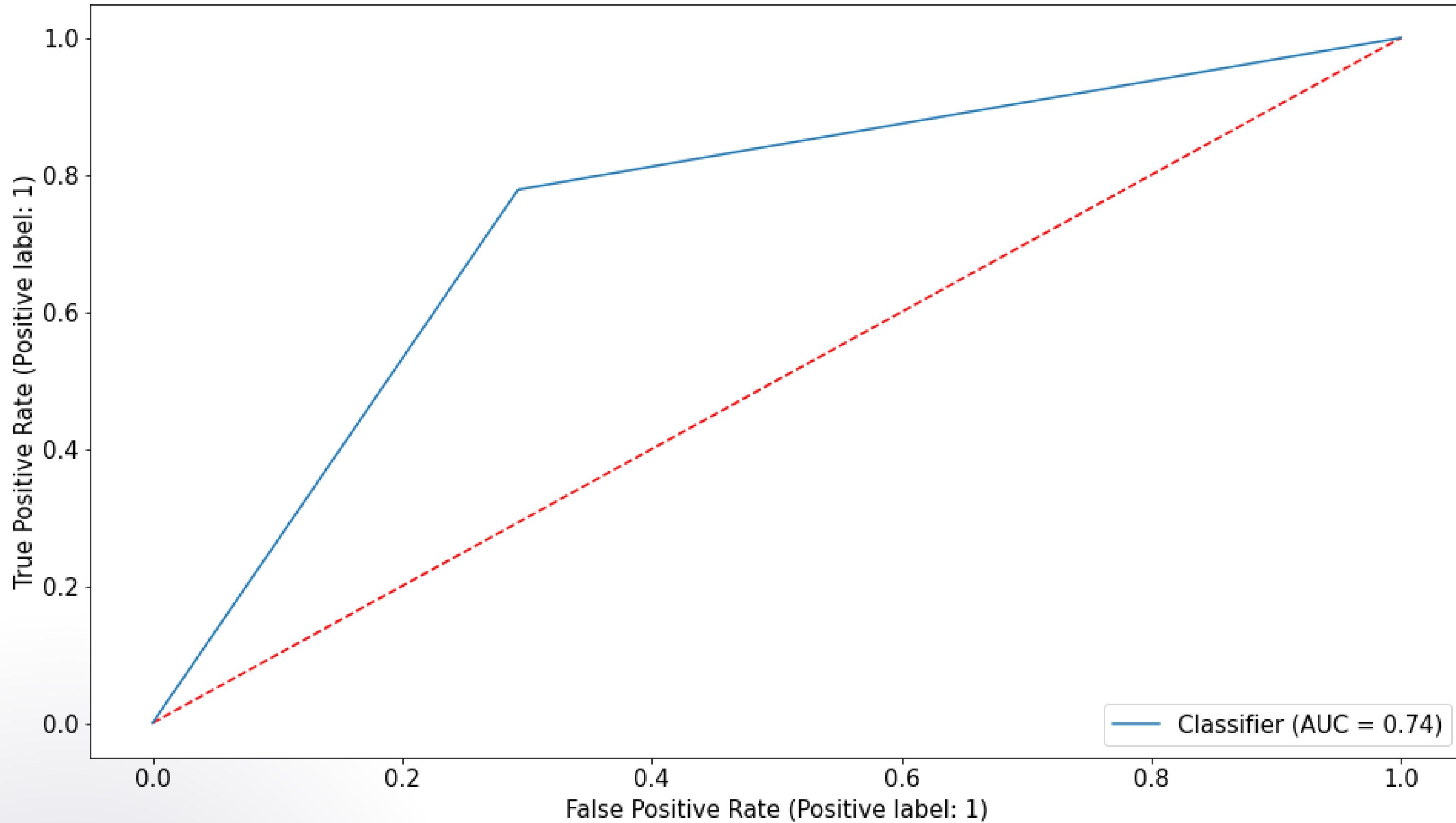
Top Combined Subreddit Words Used



XGBoost Confusion Matrix



XGBoost ROC Curve



Top 20 XGBoost Classifier Words

