

ADme Group

Bekzod Tolipov | Data Scientist





Problem Statement

Build models to identify post which subreddit it belongs and create analysis on words with high engagement posts to provide recommendations to create interactive gamified surveys to gather behavioral data which later on will be added to models to make prediction which advertisement will work on users.



EDA & NLP

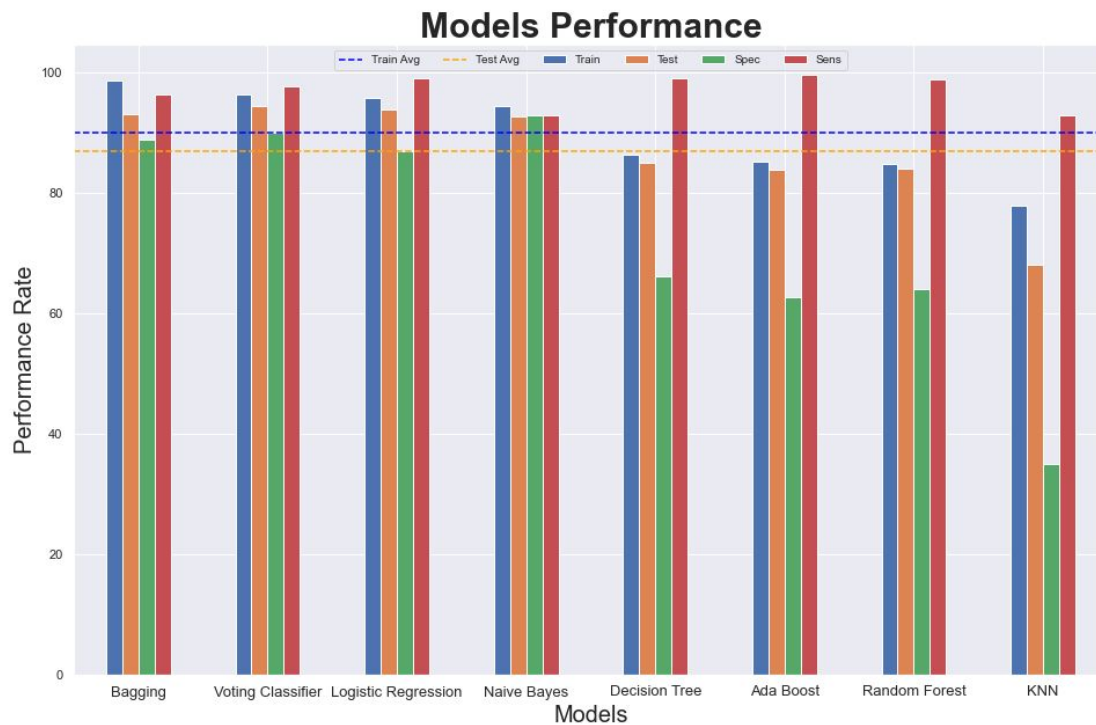
- 12,895 Records
- Baseline accuracy: 57%
- EDA:
 - Combine Title and Body text
 - Count emoji on a text
 - Regex to remove “links, png, amp, x200b”
- NLP
 - CountVectorizer | TfidfVectorizer
 - Stopwords
 - Stemmer
 - Ngram: (1,1) (2,2)
 - Sentiment Intensity Analyzer



Model Performance

Voting Classifier:

- Bagging
- Logistic Regression
- Naive Bayes

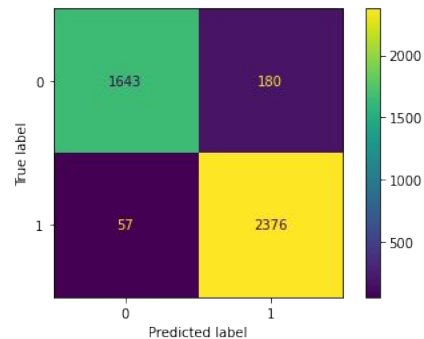




Model Performance

| | precision | recall | f1-score |
|------------------|-----------|--------|----------|
| Board Games | 0.94 | 0.98 | 0.96 |
| Wall Street Bets | 0.97 | 0.92 | 0.94 |
| accuracy | | | 0.95 |

- Specificity: 0.90
- Confidence Interval: 95.34 ± 0.01

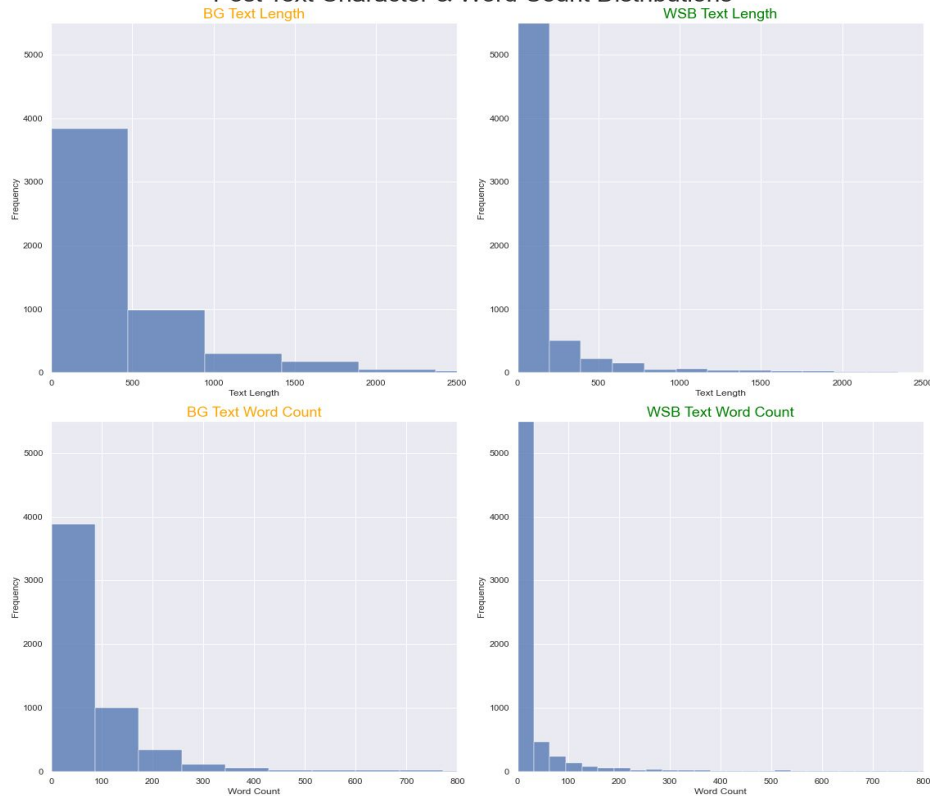




Overall Post Analysis

- **Board Games** subreddit users post tend to be more expressive and include more words 116 on average compare to **Wall Street Bets** 37 average

Post Text Character & Word Count Distributions

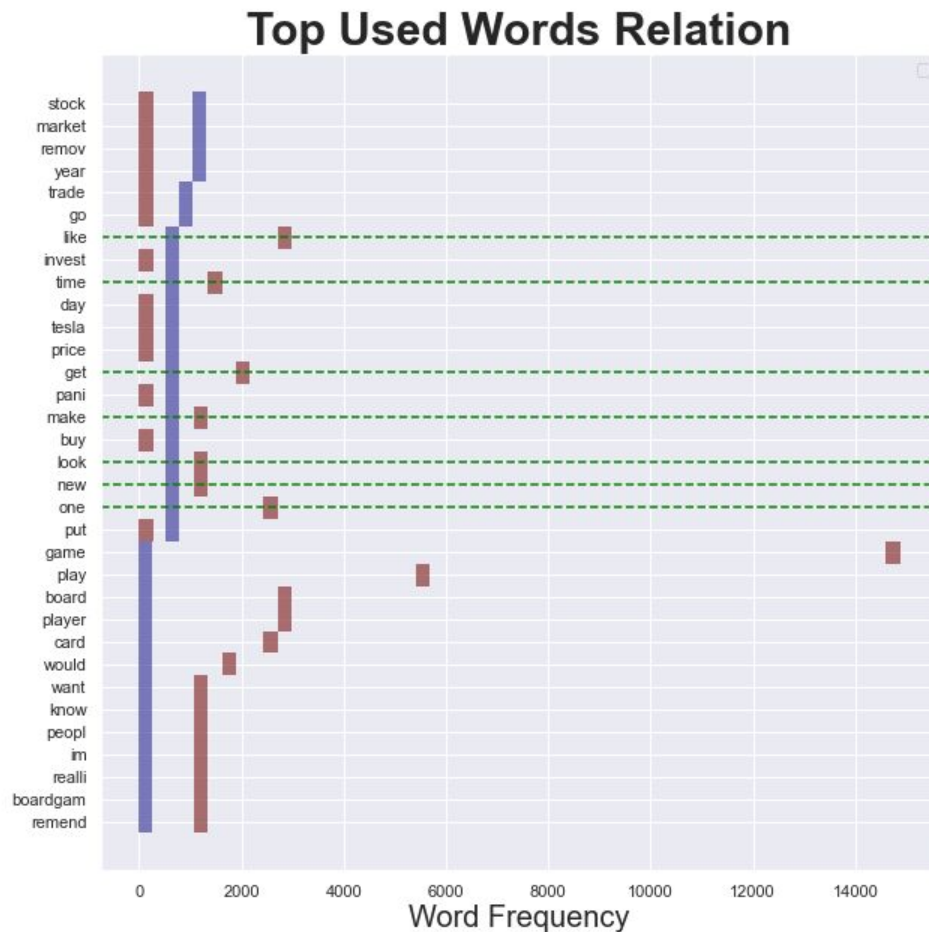




Subreddit comparison

Common words from top 20 words:

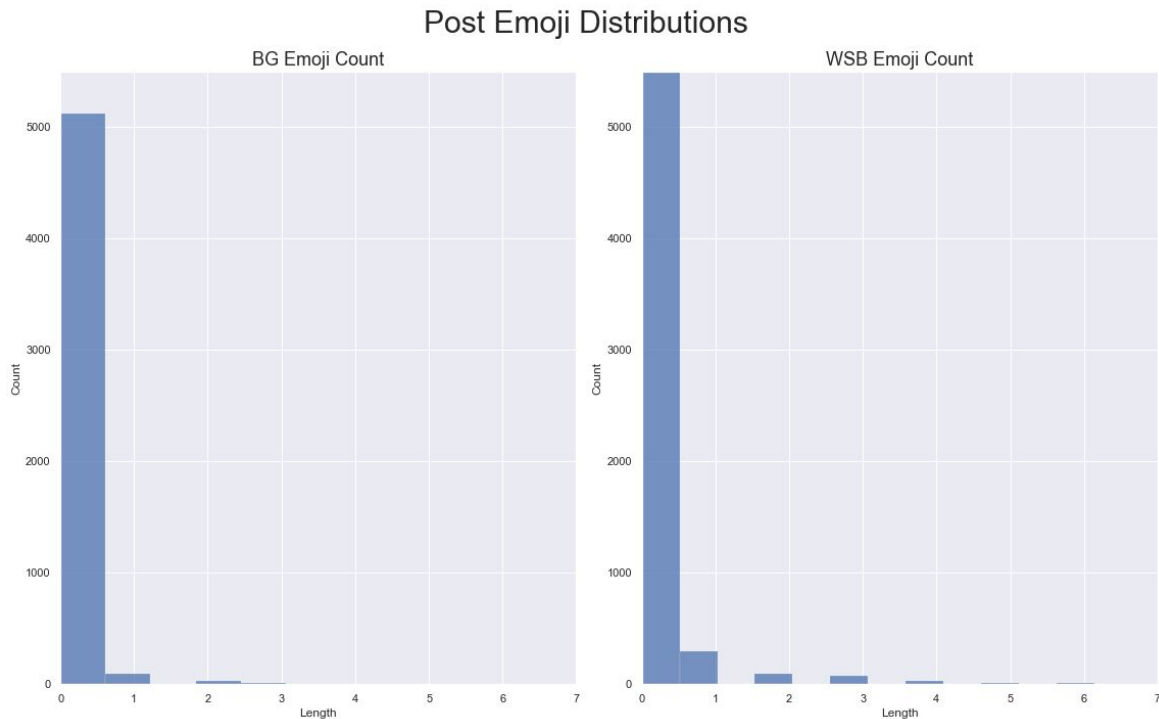
1. Like
2. Time
3. Get
4. Make
5. Look
6. New
7. One





Subreddit emoji Comparison

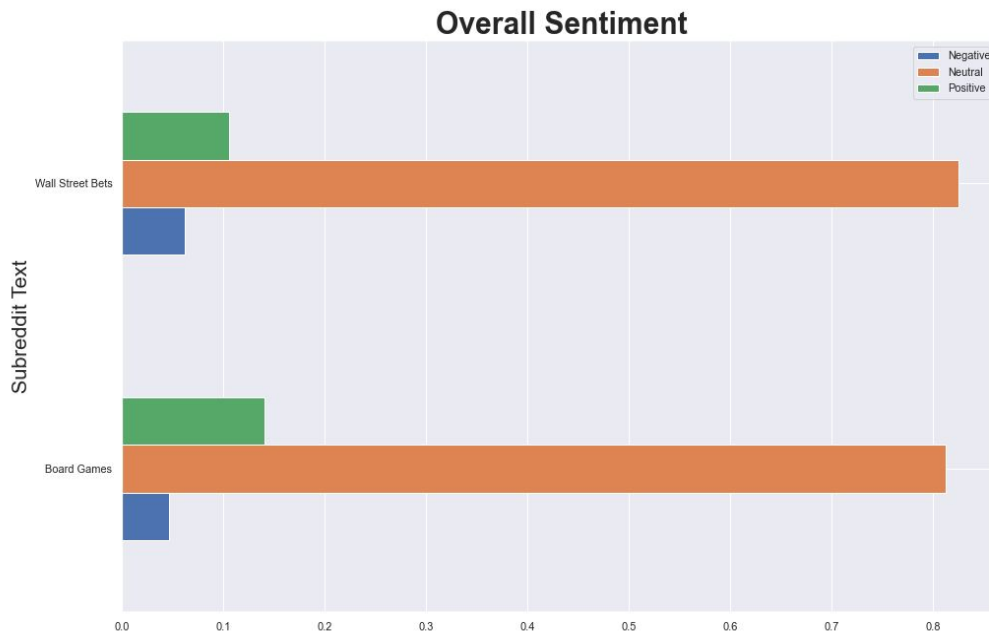
Both subreddits shows **low** activity of emoji use





Subreddit Sentiment Comparison

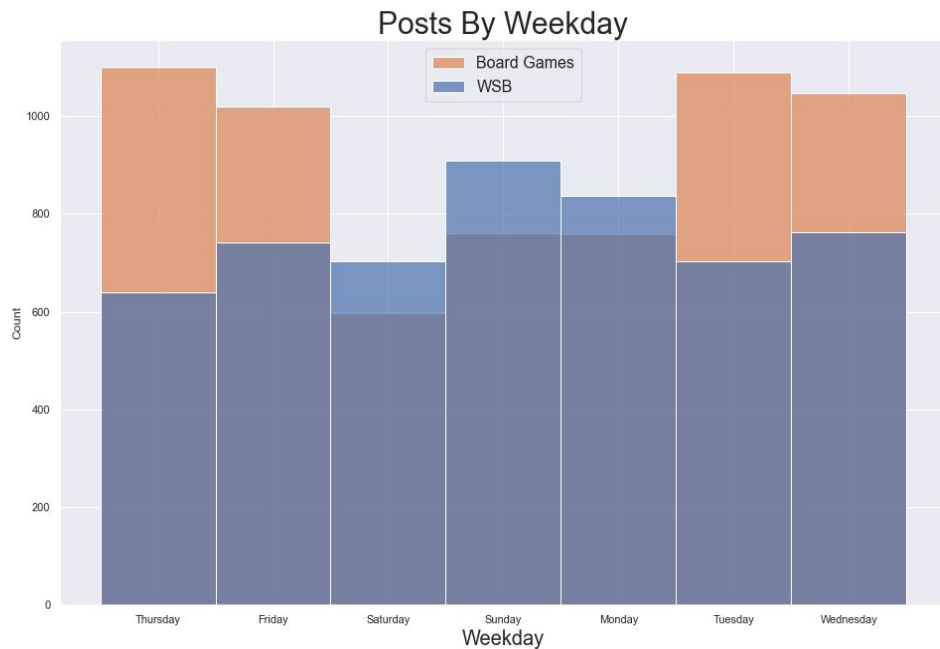
Board Games leans towards positive posts compare to Wall Street Bets but overall neutral





Subreddit Post By Weekday Comparison

Board Games more active during weekdays and less on weekends compare to Wall Street Bets



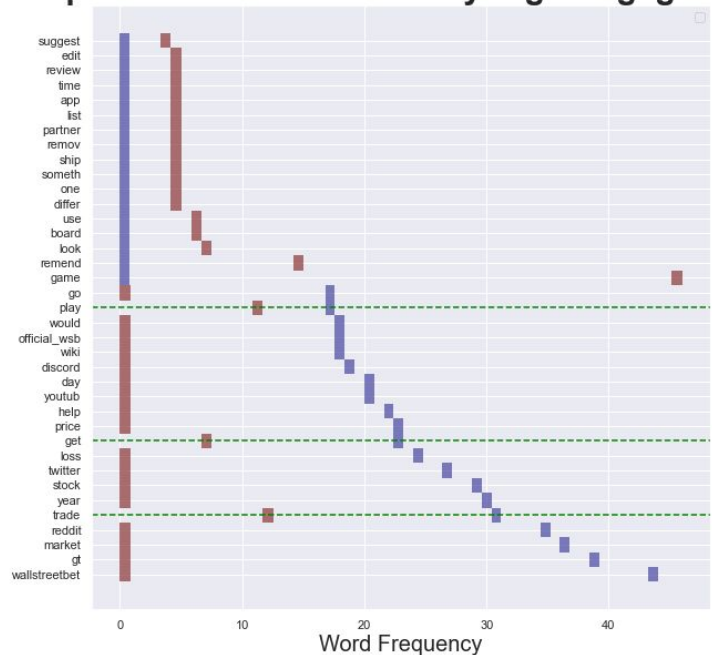


Highest Engagement Posts

Common Words:

1. Play
2. Get
3. Trade

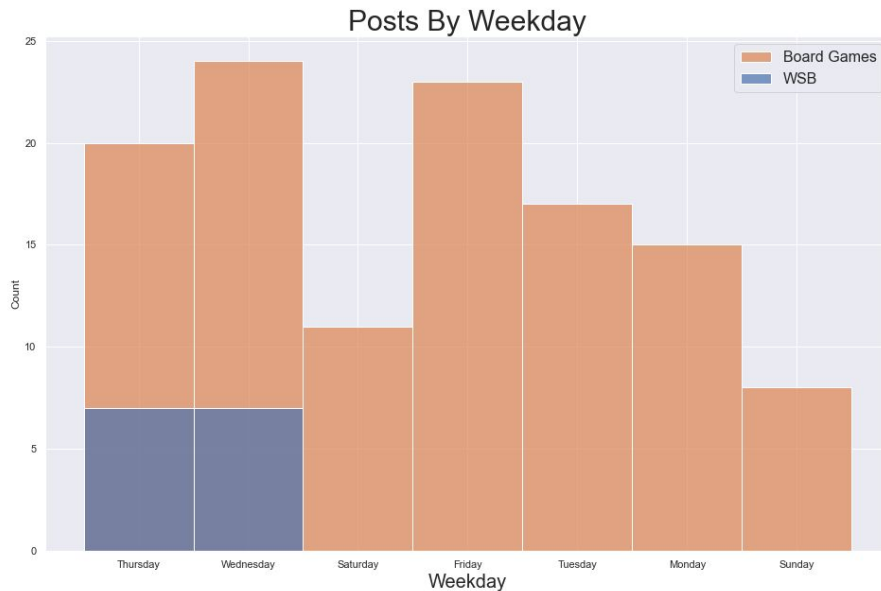
Top Used Words Relation By High Engagement





Highest Engagement Posts

Board Games highest engagement posts happen during Wednesday and Friday. Wall Street Bets on Wednesday and Thursday





Recommendations

Our recommendation includes:

- Use following words to create interactive survey to collect behavioral scoring from both subreddits:
 - ['market', 'reddit', 'year', 'stock', 'twitter', 'loss', 'price', 'get', 'help', 'day', 'youtube', 'discord', 'would', 'wiki', 'go', 'play']
- Post it on Wednesday, Thursday and Friday
- Survey topic ['trade']