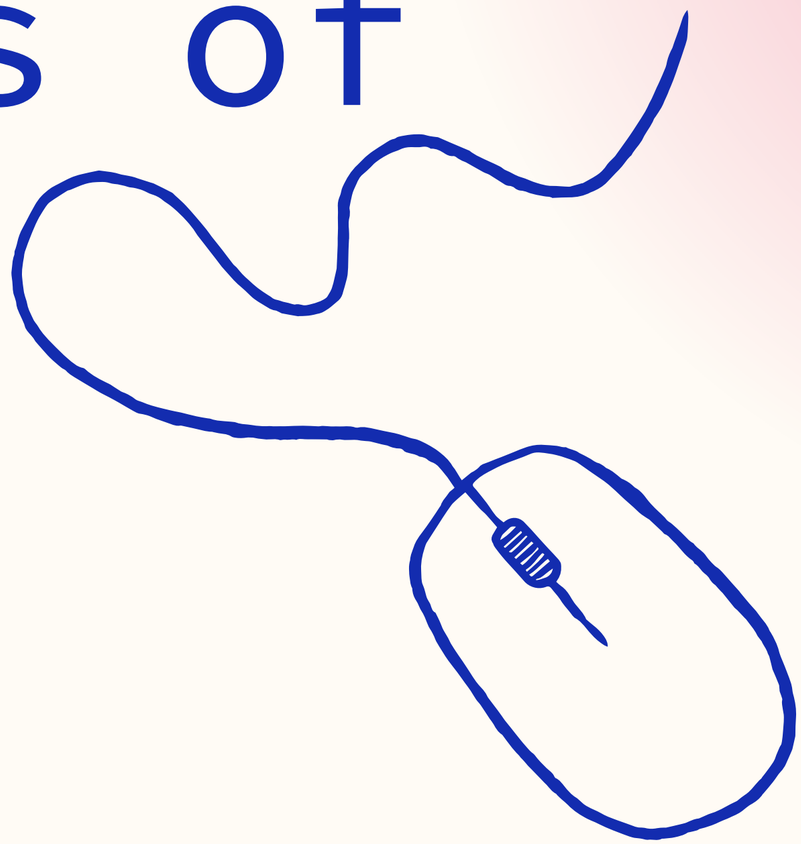
Extensive Sentiment Analysis of Tweets



Ram Terli
Siddharth Maredu
Srinivas Yedla

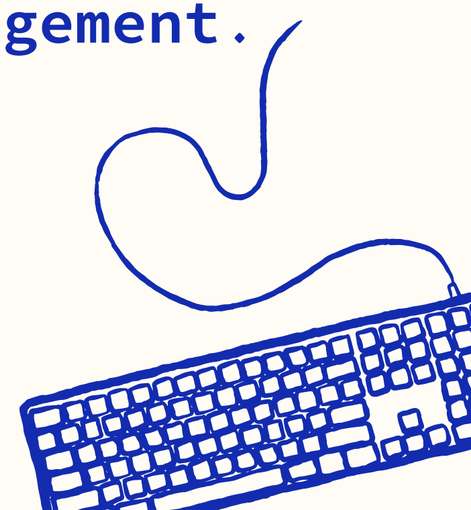
NO MORE
HATE SPEECH

Project Proposal:

"How can identifying hate speech in tweets help businesses better manage their brand reputation online?"

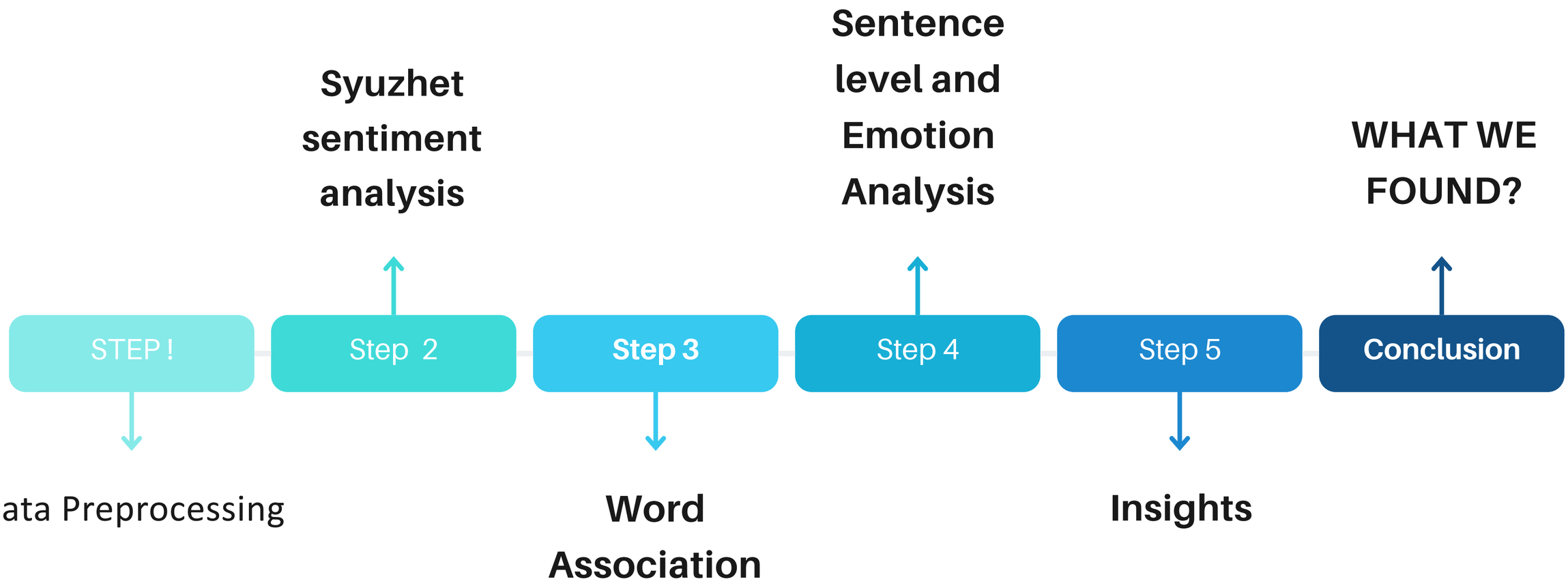
Importance:

By identifying and mitigating hate speech, Twitter can create a safer and more positive user environment, thereby protecting its brand reputation and user engagement.



THE PROJECT DESIGN

A brief history of what we did



01 - Data Preprocessing

Basic Cleaning-removed user mentions, URLs, hashtags, special characters, and extra whitespaces. It also converts text to lowercase, expands contractions, corrects common slang, and handles emojis.

Removing Stop Words-The tweets are tokenized into words, and stopwords (common words like "the", "and", etc.) are removed.
--The length of each cleaned tweet is calculated, and non-finite values are handled.
.

Syuzhet sentiment analysis

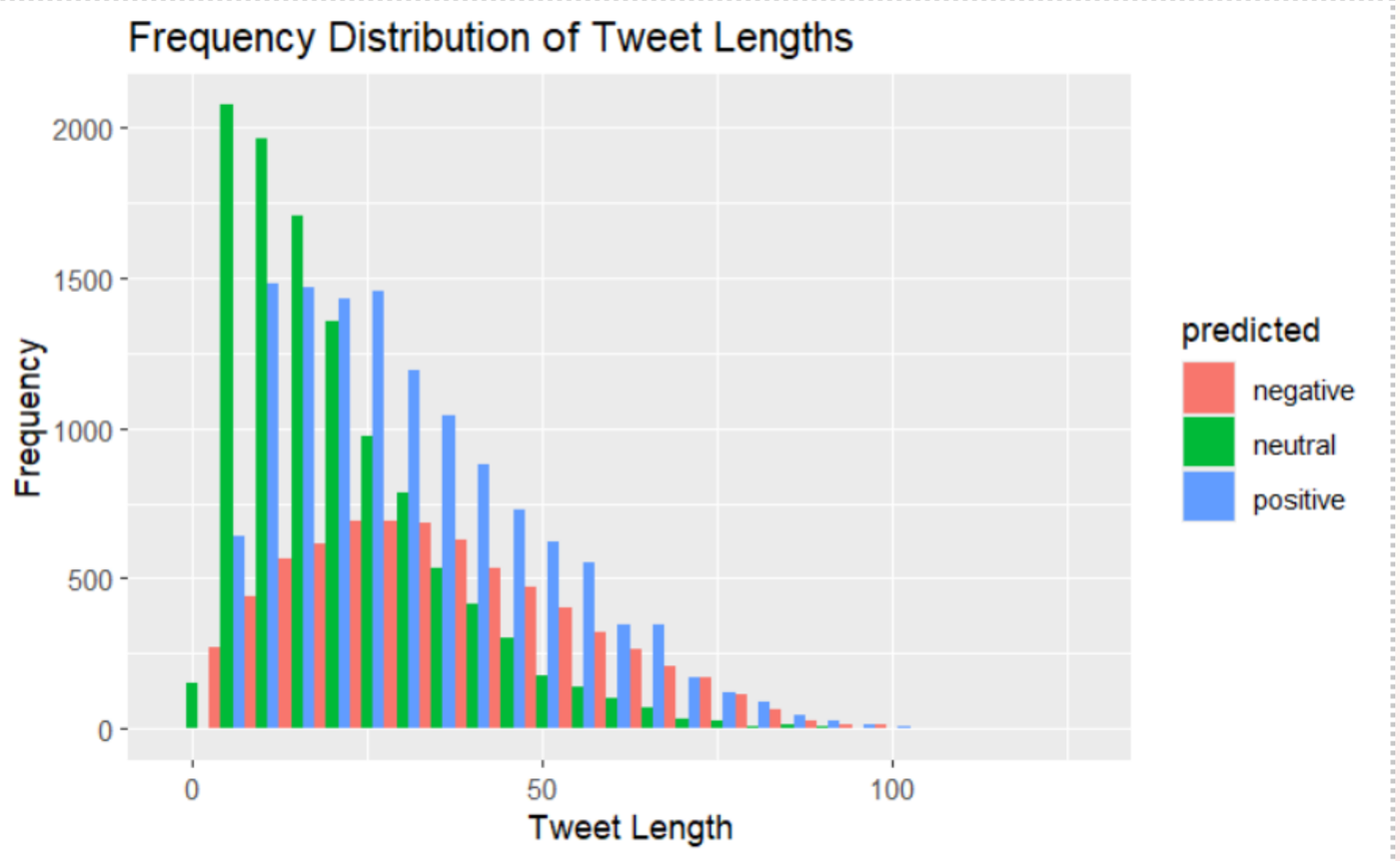
- The syuzhet method uses the NRC Emotion Lexicon to calculate sentiment scores.
- It assigns a numeric sentiment score to each tweet based on the presence of positive and negative words.

A function classify_sentiment is defined to classify each tweet as positive, negative, or neutral based on its sentiment score using the syuzhet package.

Word Clouds

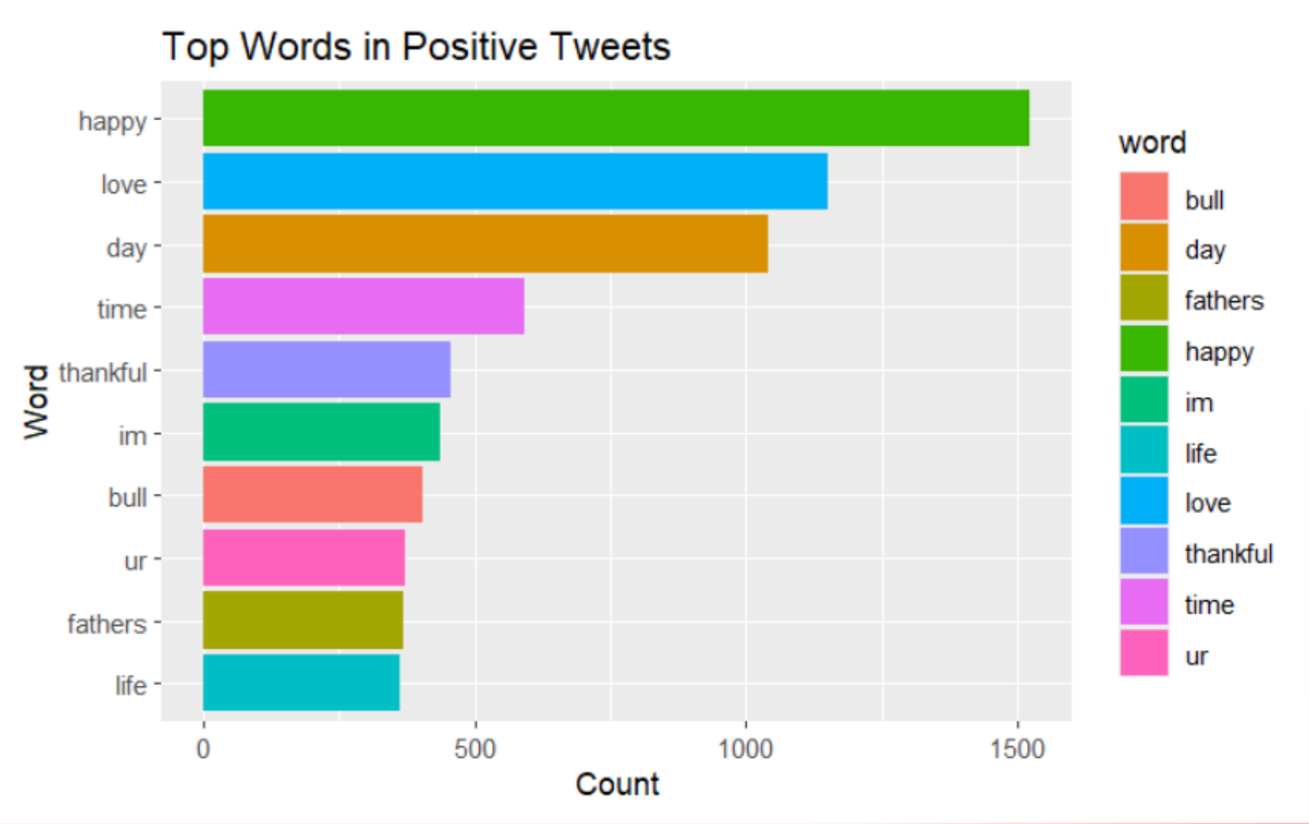
Tweets are filtered into positive, negative, and neutral based on their predicted sentiment.

Frequency Distribution of Tweet Lengths

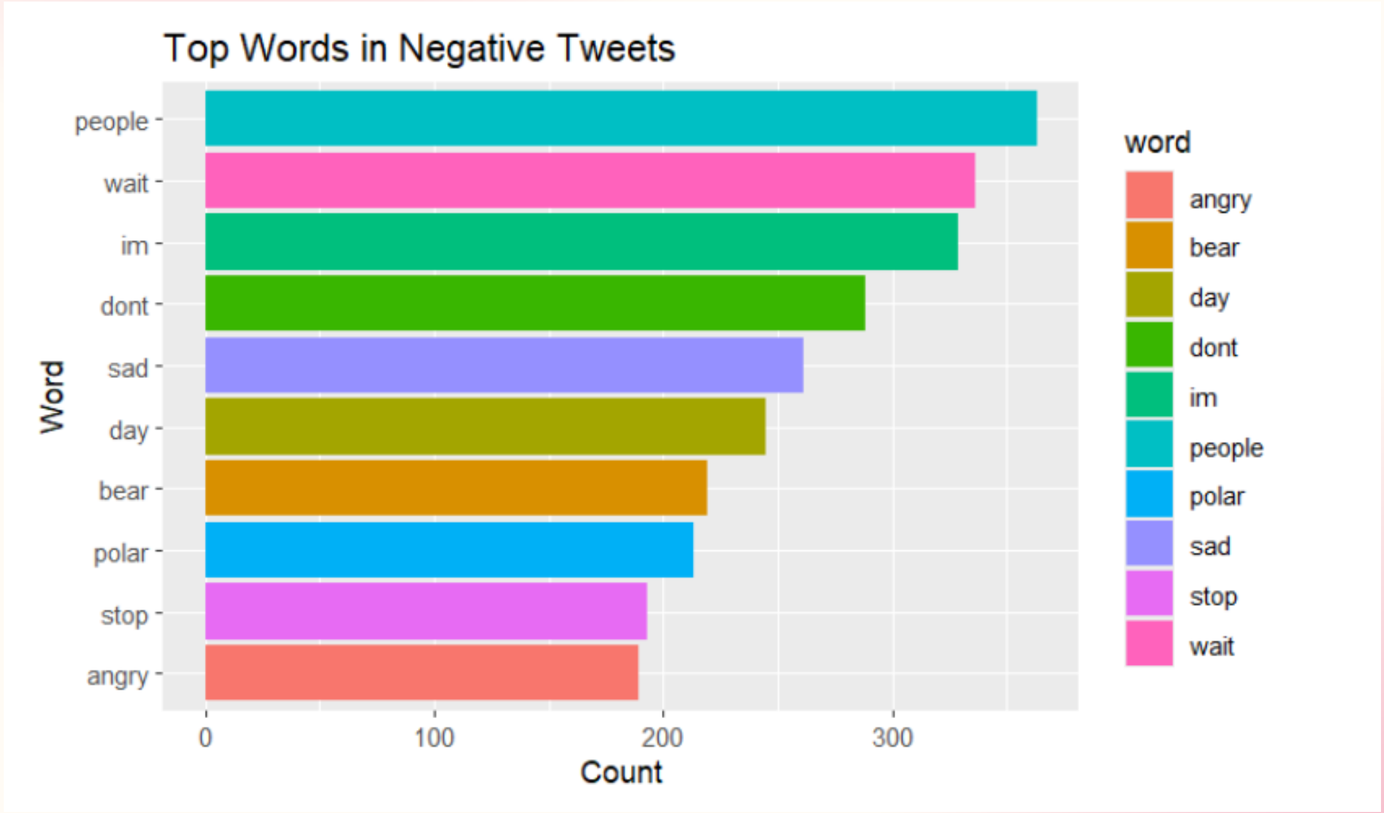


This plot provides us the length of each type of tweet classified by sentiment. This gives us an insight about how each tweet can be possibly classified into just based on length.

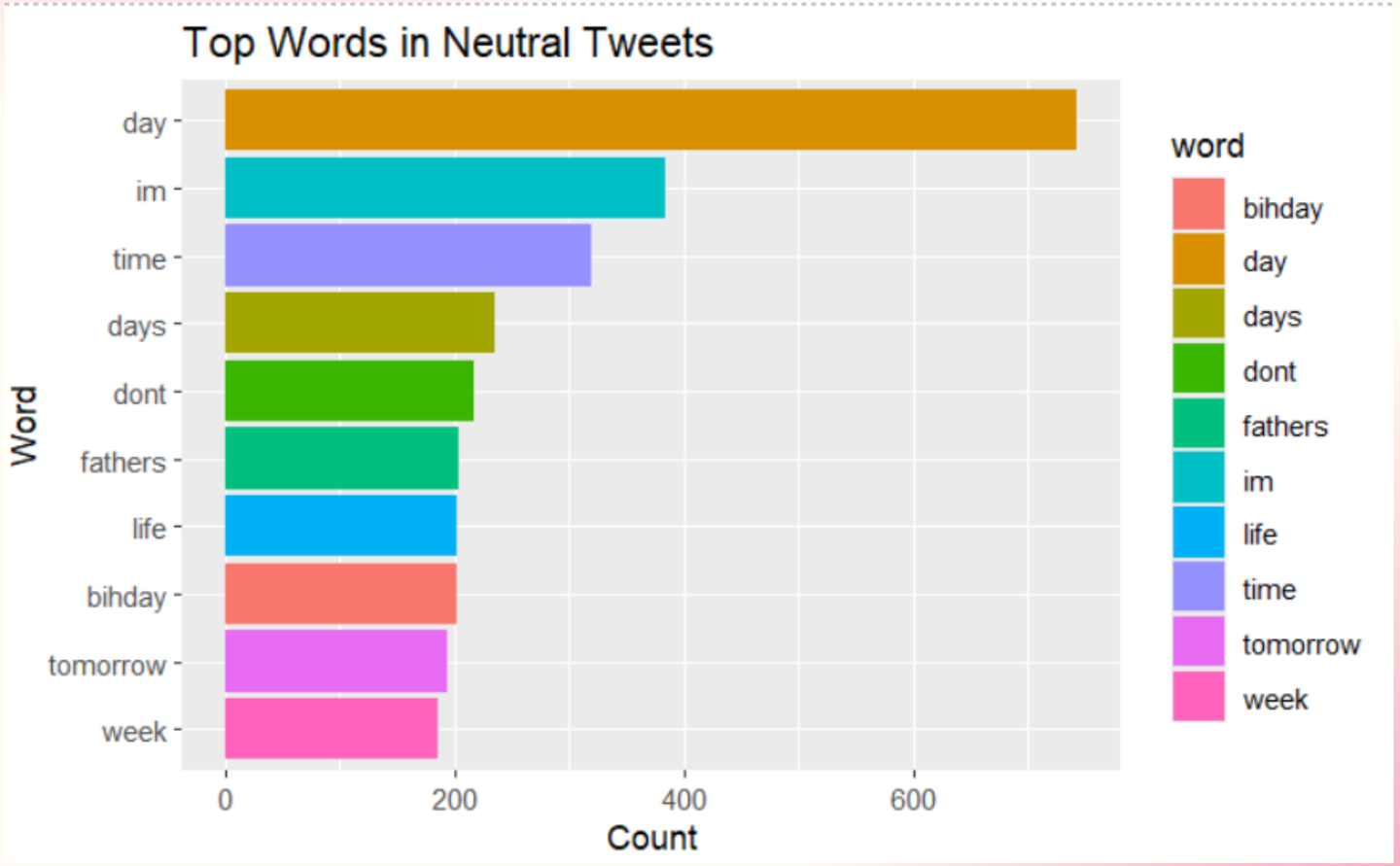
TOP WORD COUNTS IN POSITIVE TWEETS



TOP WORD COUNTS IN NEGATIVE TWEETS



TOP WORD COUNTS IN NEUTRAL TWEETS



WORD ASSOCIATION BY SENTIMENT: Correlation score

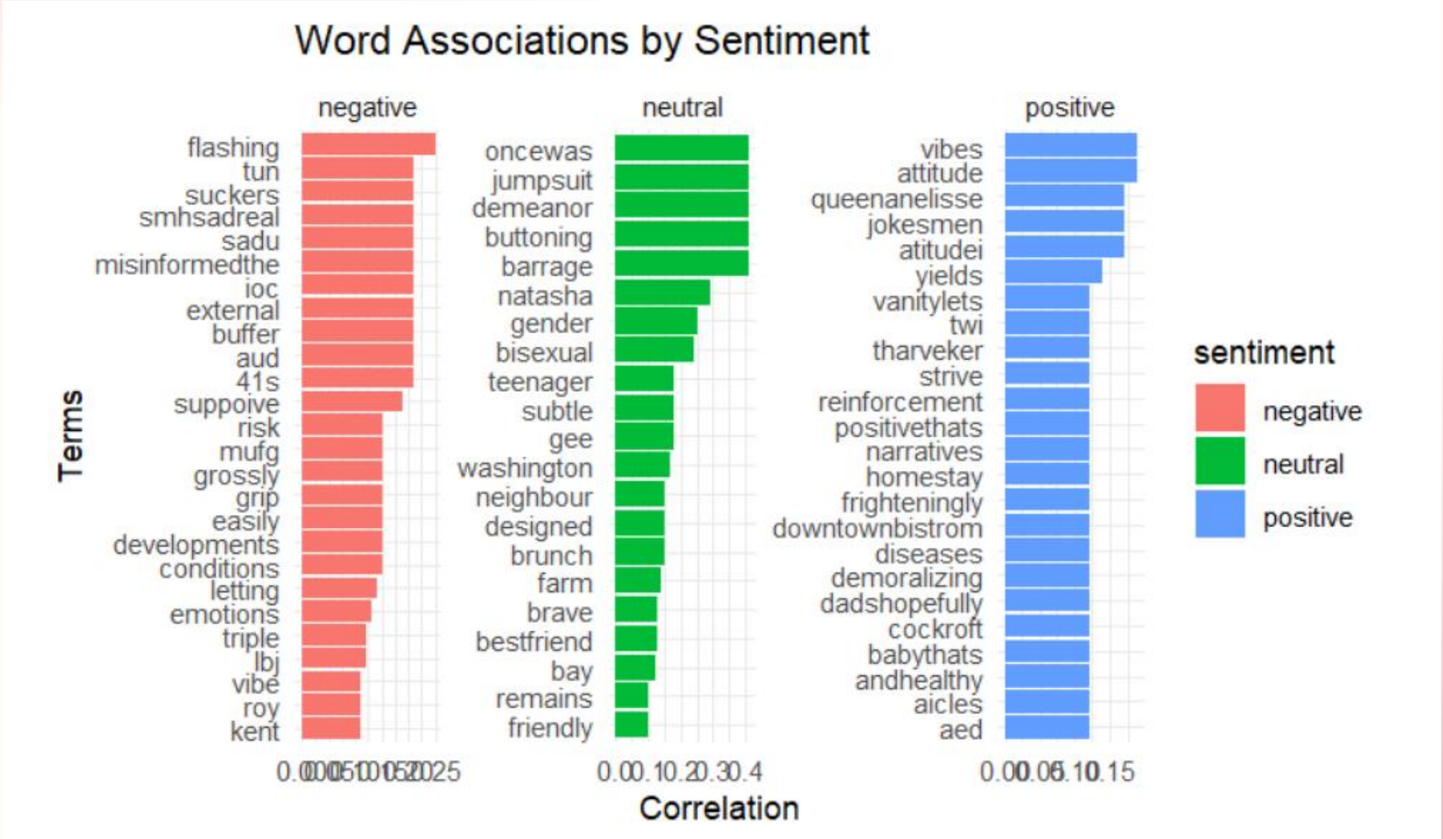
```
## $positive
##           vibes        attitude        atitudei        jokesmen   queenanelisse
##            0.19            0.19            0.17            0.17            0.17
##          yields             aed      babythats        diseases downtownbistrom
##            0.14            0.12            0.12            0.12            0.12
##             twi       tharveker   dadshopefully        homestay          aicles
##            0.12            0.12            0.12            0.12            0.12
##      vanitylets      narratives   frighteningly      andhealthy   positivethats
##            0.12            0.12            0.12            0.12            0.12
##    demoralizing        cockroft   reinforcement          strive
##            0.12            0.12            0.12            0.12
##
## $negative
##       flashing        external         suckers             aud          buffer
##           0.25            0.21            0.21            0.21            0.21
## misinformedthe             tun             ioc             41s            sadu
```

WORD ASSOCIATION BY SENTIMENT: Correlation score

```
##          0.21          0.21          0.21          0.21          0.21
##     smhsadreal       suppoive       easily          risk          mufg
##          0.21          0.19          0.15          0.15          0.15
##          grip   developments    conditions       grossly       letting
##          0.15          0.15          0.15          0.15          0.14
##      emotions        triple           lbj          kent           roy
##          0.13          0.12          0.12          0.11          0.11
##          vibe
##          0.11
##
## $neutral
##    barrage   demeanor    oncewas  buttoning   jumpsuit    natasha     gender
##       0.41       0.41       0.41       0.41       0.41       0.29       0.25
##   bisexual     subtle        gee   teenager washington     brunch  neighbour
##       0.24       0.18       0.18       0.18       0.17       0.15       0.15
##   designed       farm      brave bestfriend        bay    remains   friendly
##       0.15       0.14       0.13       0.13       0.12       0.10       0.10
```

visualize associations

WORD ASSOCIATION BY SENTIMENT



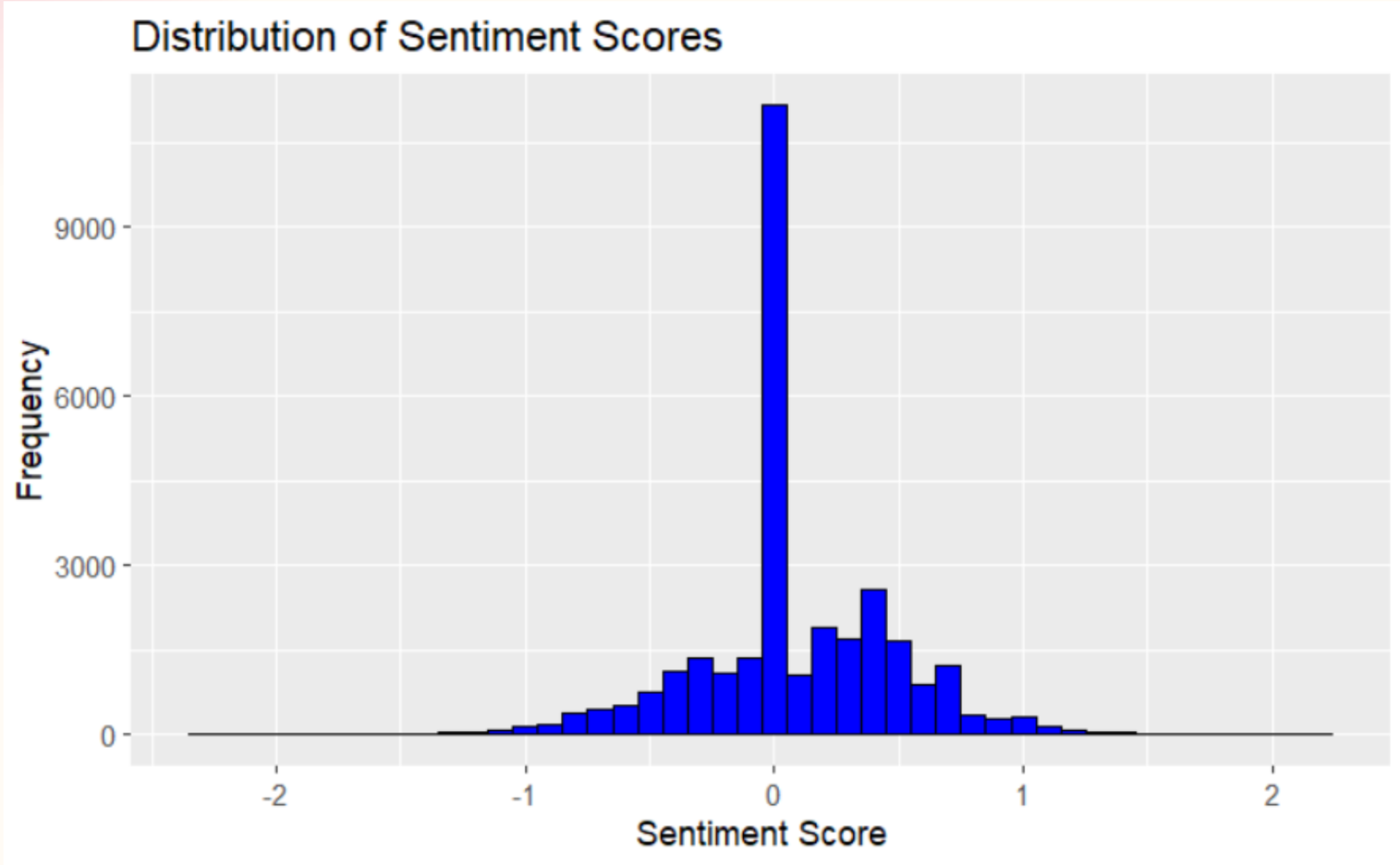
WORD ASSOCIATION BY SENTIMENT: Word clouds

SENTIMENTS USING SENTENCE LEVEL ANALYSIS

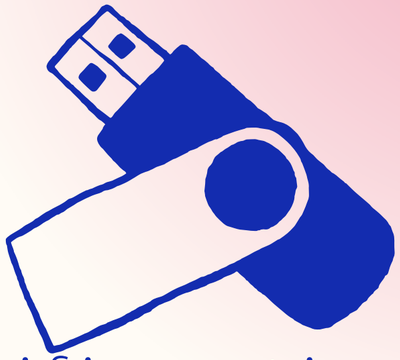
The sentimentr package is designed to analyze sentiment at the sentence level, which allows for more nuanced sentiment analysis compared to word-level methods.

SENTIMENT SCORE DISTRIBUTION

negative	neutral	positive
7576	10628	12476



EMOTIONAL ANALYSIS



get_nrc_sentiment Function: calculates the presence of specific emotions and sentiments in each tweet using the NRC Emotion Lexicon

nrc_sentiment Dataframe: returns a dataframe where each row corresponds to a tweet, and each column represents the count of words associated with a particular emotion or sentiment.

anger	anticipation	disgust	fear	joy	sadness
5083	10855	3790	5595	11172	4869

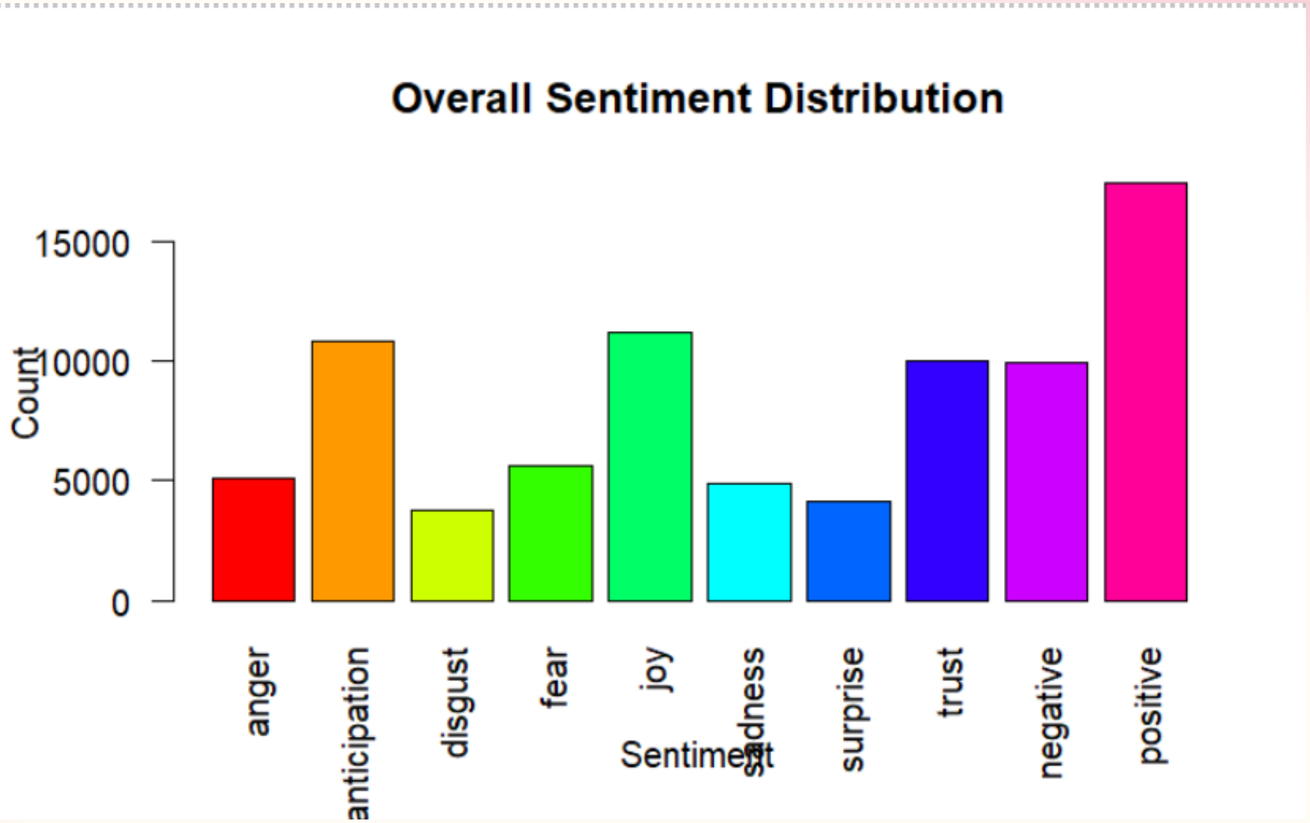
surprise	trust	negative	positive
4126	9991	9950	17445

EMOTIONAL ANALYSIS

syuzhet Method:

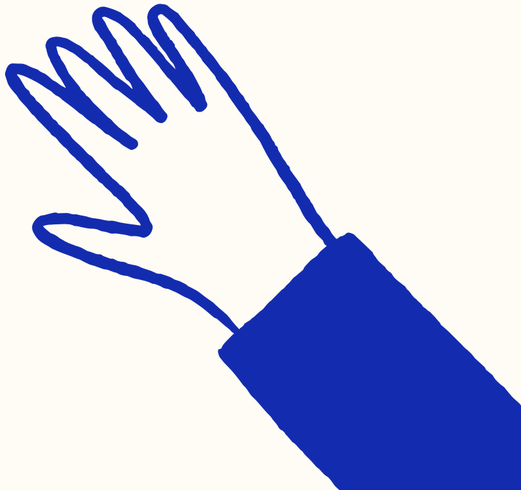
Quick and straightforward way to gauge overall sentiment.
Useful for high-level sentiment trends and polarity detection.
Limited in terms of emotional detail and specificity.
get_nrc_sentiment Method:

Provides a comprehensive emotional profile.
Useful for detailed emotional analysis, helping to understand not just whether the sentiment is positive or negative, but also the specific emotions driving that sentiment.
More complex output that requires more detailed analysis but provides richer insights.



INSIGHTS

Histogram of train_data$label
Frequency
0
10000
25000
0.0
0.2
0.4
0.6
0.8
1.0
train_data$label



Conclusion

Robust analysis to identify sentiments of tweets.
Valuable insights for twitter business to handle ethical implications of their platform
Awareness for Investors

Thanks

Any
Questions?