SMA Group Project Spotify Analysis

ROSATO Alessio BAJPAI Shruti YANG Zhou

Feb 03, 2021

```
## package 'RWeka' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\ziyth\AppData\Local\Temp\RtmpIfp4Bp\downloaded_packages
```

Overview of the report

1.About Company

Spotify is a genuine streaming media music service platform, which was officially launched in Stockholm in October 2008.

2.Project Description

To search through Twitter API to obtain a large number of data information related to spotify (including published content, comments, sources, user information, etc.), and to classify and compare the source data through sentiment analysis and topic analysis, so as to obtain the participation and liking degree of spotify users on different emotions and topics.

3.Data Description

spotify_today : Raw data from twitter. Spotify_tweets_organic : Original twitter data filtered from original data. Spotify_retweets : Retweet data filtered from raw data. Spotify_replies : reply data filtered from raw data.
working_tweets: Volume over time (last 3200 posts by the official account, approx. 9 months)

4.Data Pre-processing

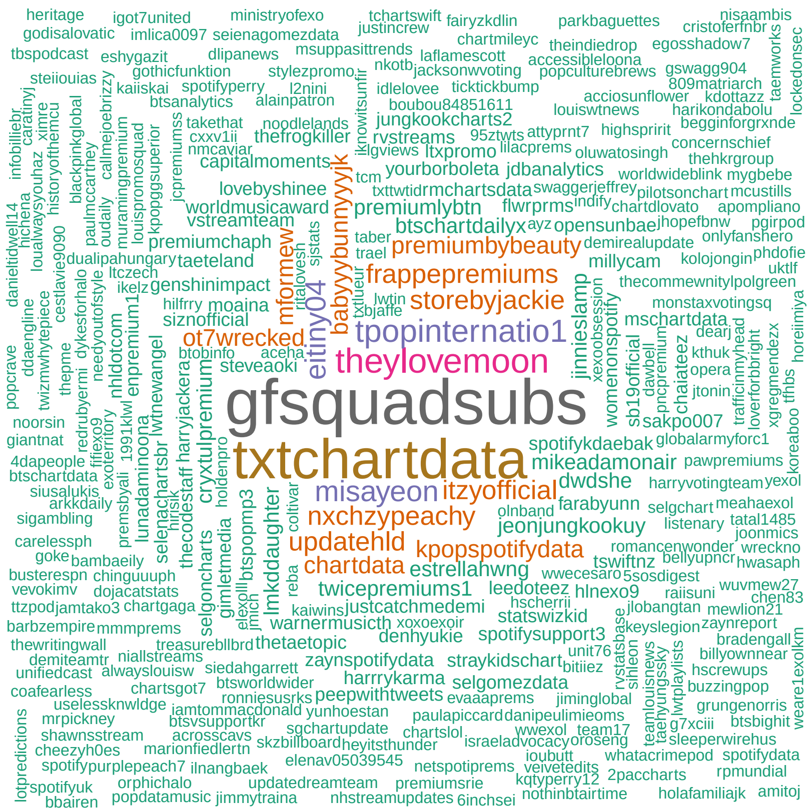
The preprocessing of twitter text is helpful for our data analysis in the future. First,We remove punctuation and numbers Second, convert the text to lowercase Third, remove stop words Finally, the source data is divided into Spotify_tweets_organic, Spotify_retweets and Spotify_replies

5.User distribution

```
## # A tibble: 5 x 2
##   place_full_name               n
##   <chr>                     <int>
## 1 Batangas City, Calabarzon    54
## 2 Los Angeles, CA              46
## 3 Brooklyn, NY                 44
## 4 Orlando, FL                  28
## 5 Laredo, TX                   26
```

The above data shows the six places with the largest number of users in the dataset.According to the distribution of users, we can know which regions are more concerned about spotify's Twitter, and which regions are less concerned about spotify's Twitter. The company can modify its marketing strategy more specifically, so as to gain greater popularity and more users.

6.Users who actively forward twitter

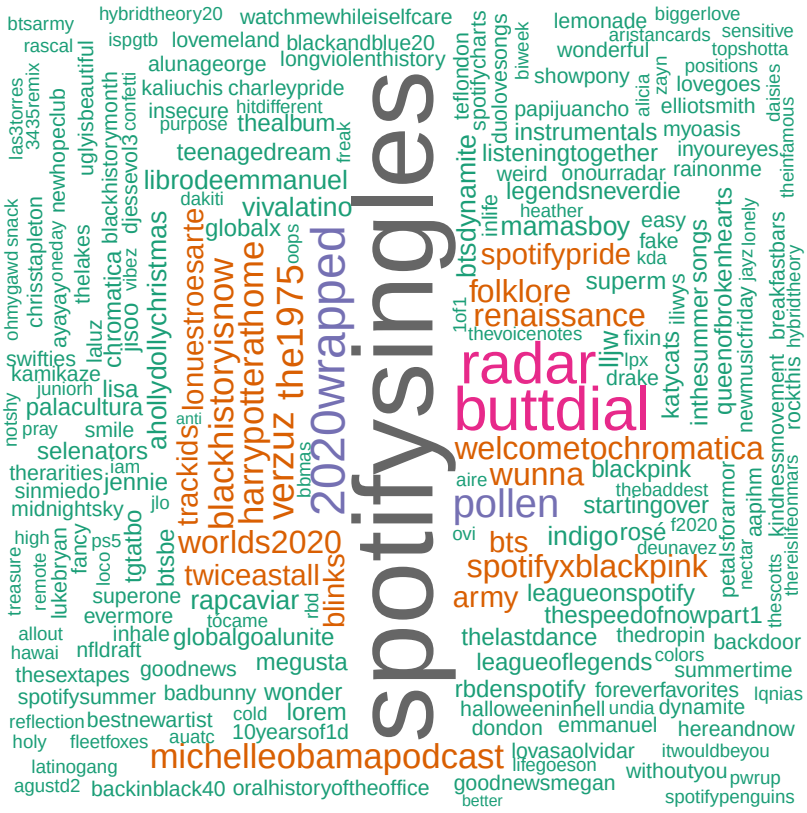


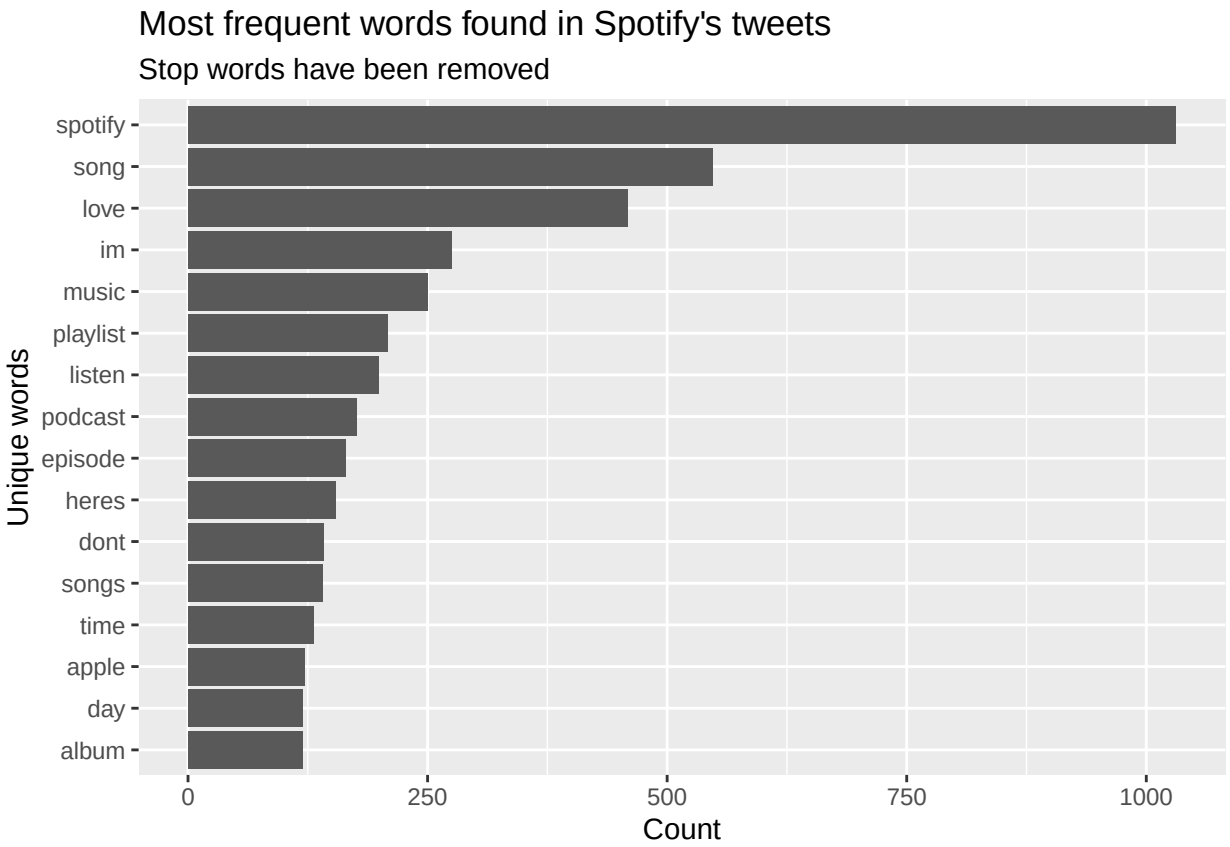
7.Users who actively forward twitter recently

Recently actively forwarding the data of Twitter users can help us find the recently active customers

Paying attention to the users (loyal users) who frequently forward the company's Twitter, their emotions on the information released by the company's Twitter are often a very important part, and keeping them in a high positive mood is an important part for the company to reduce the loss of users.

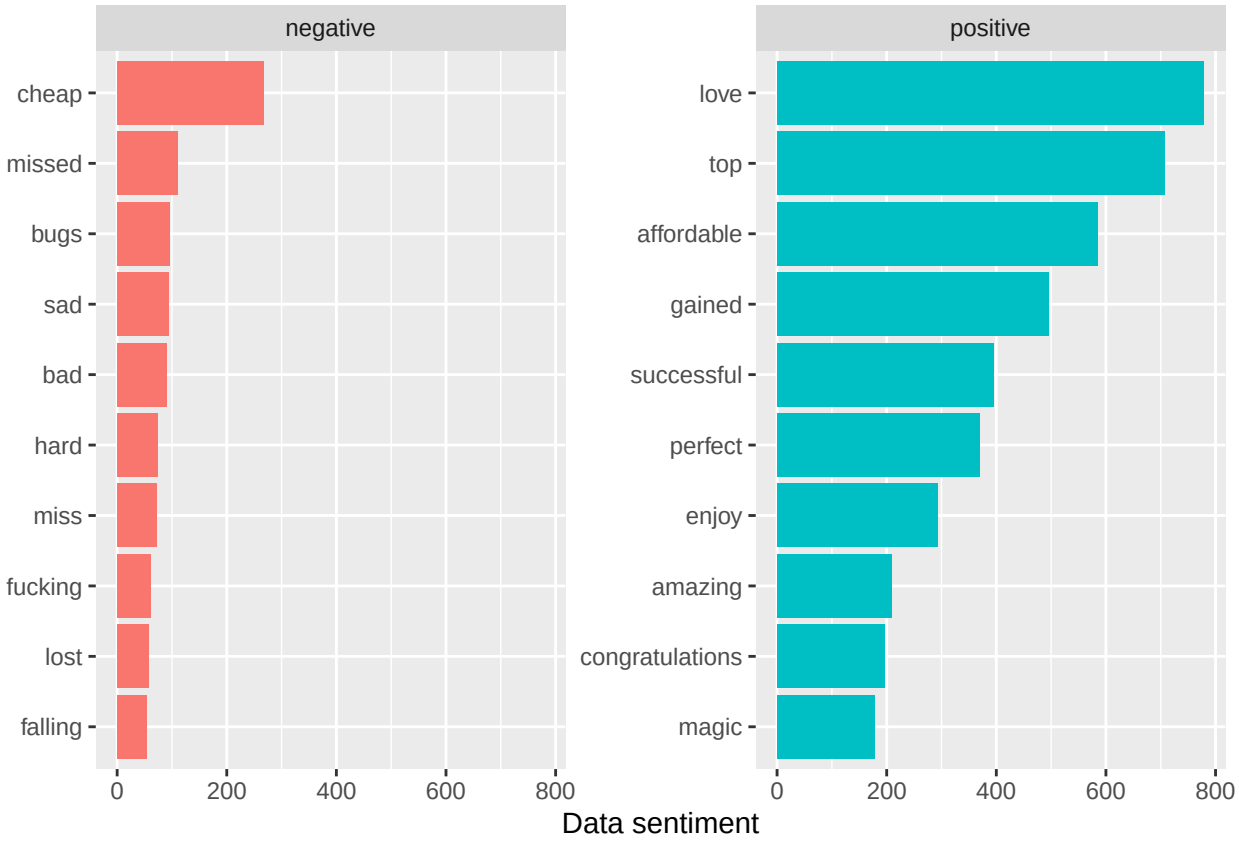
8.Most Common Hashtags

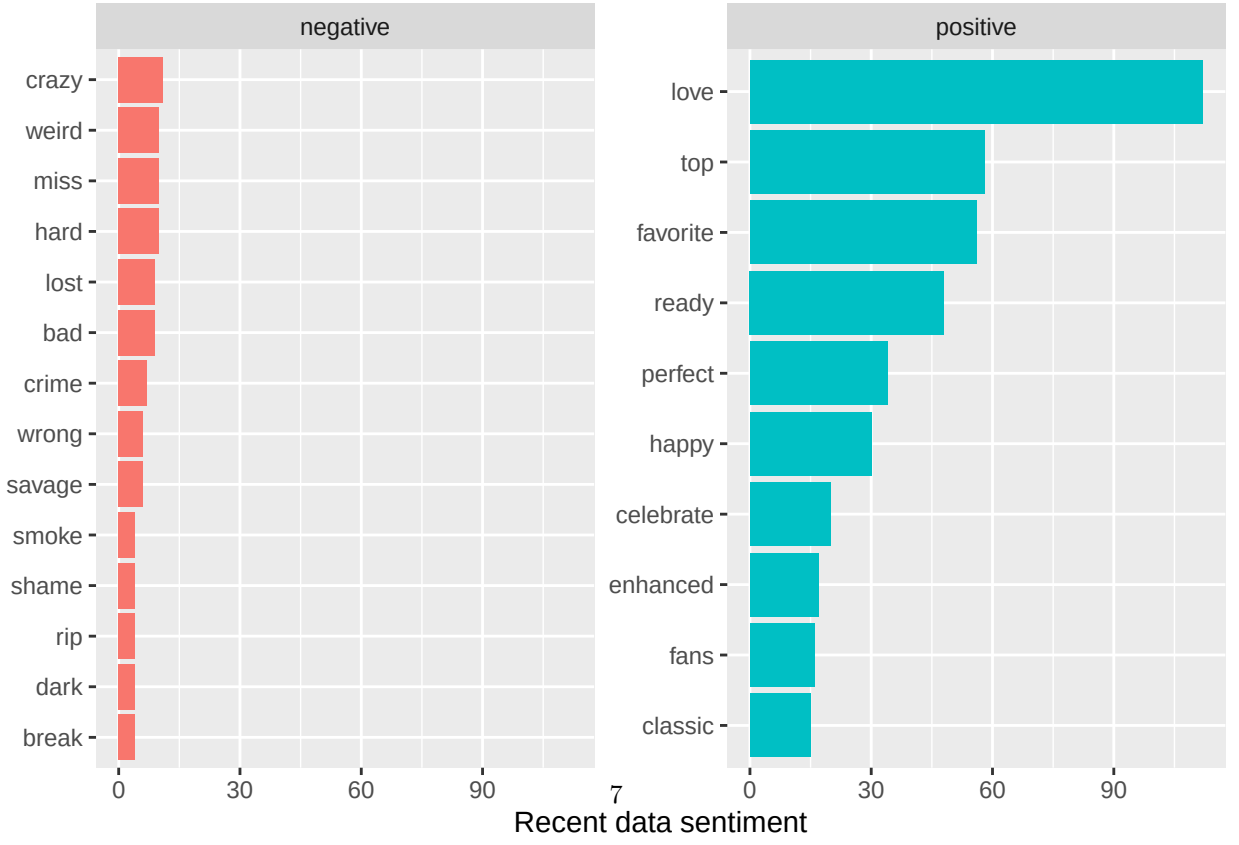



Most frequent words found in Spotify's tweets
Stop words have been removed
Unique words
spotify
song
love
im
music
playlist
listen
podcast
episode
heres
dont
songs
time
apple
day
album
0
250
500
750
1000
Count

Sentiment Analysis

Raw data Analysis



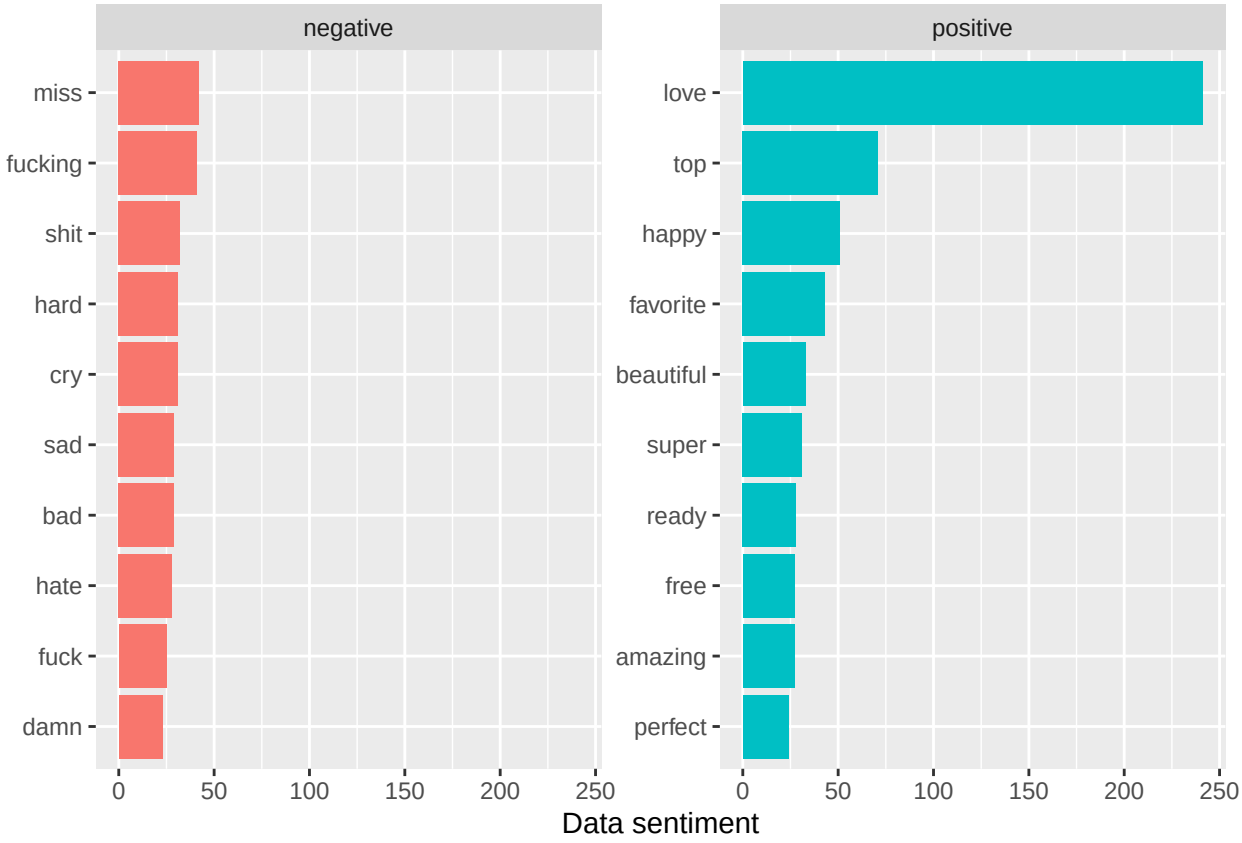


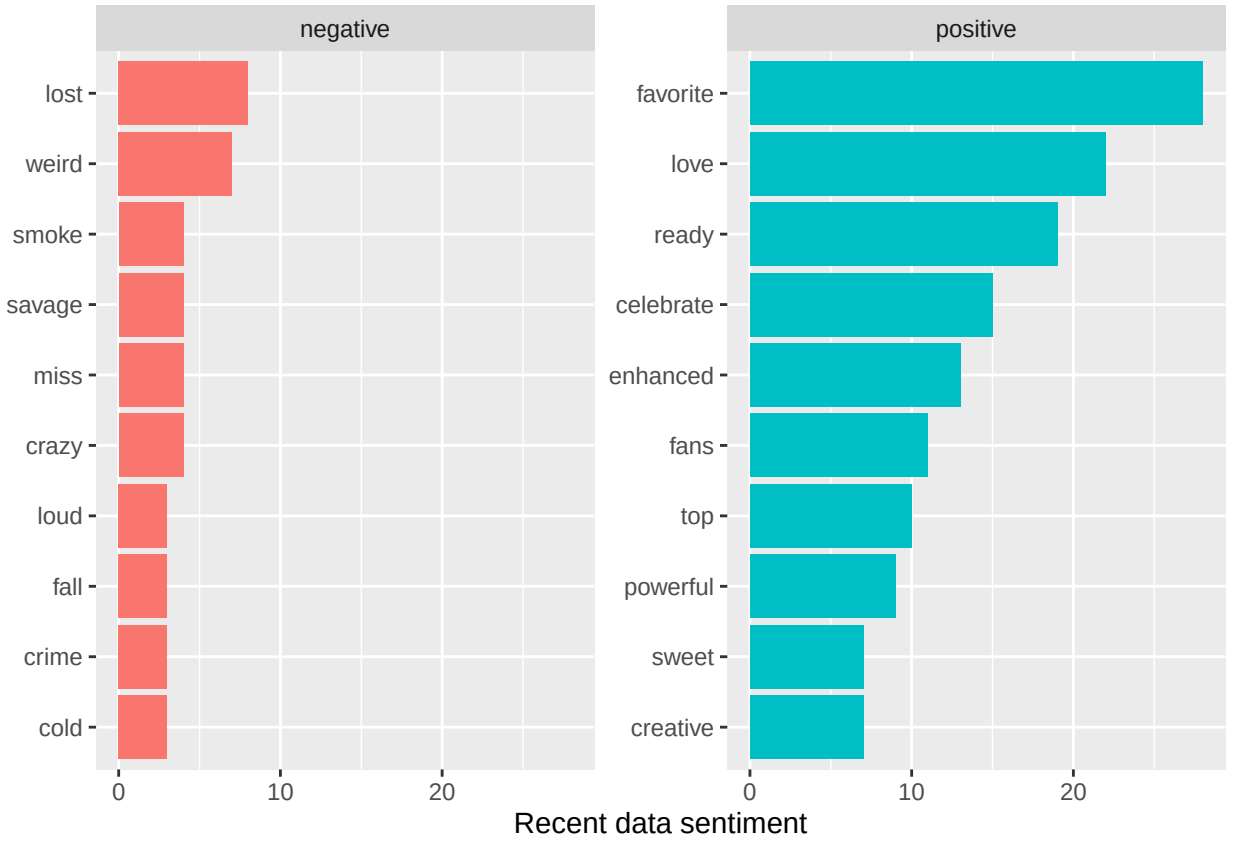
```
## [1] "Mean sentiment score = 0.920431361119207"
```

```
## [1] "Mean Working sentiment score = 0.660931174089069"
```

As can be seen from the chart, the average sentiment score of recent twitter data is higher than that of a large number of Spotify company we have chosen.

Original twitter data



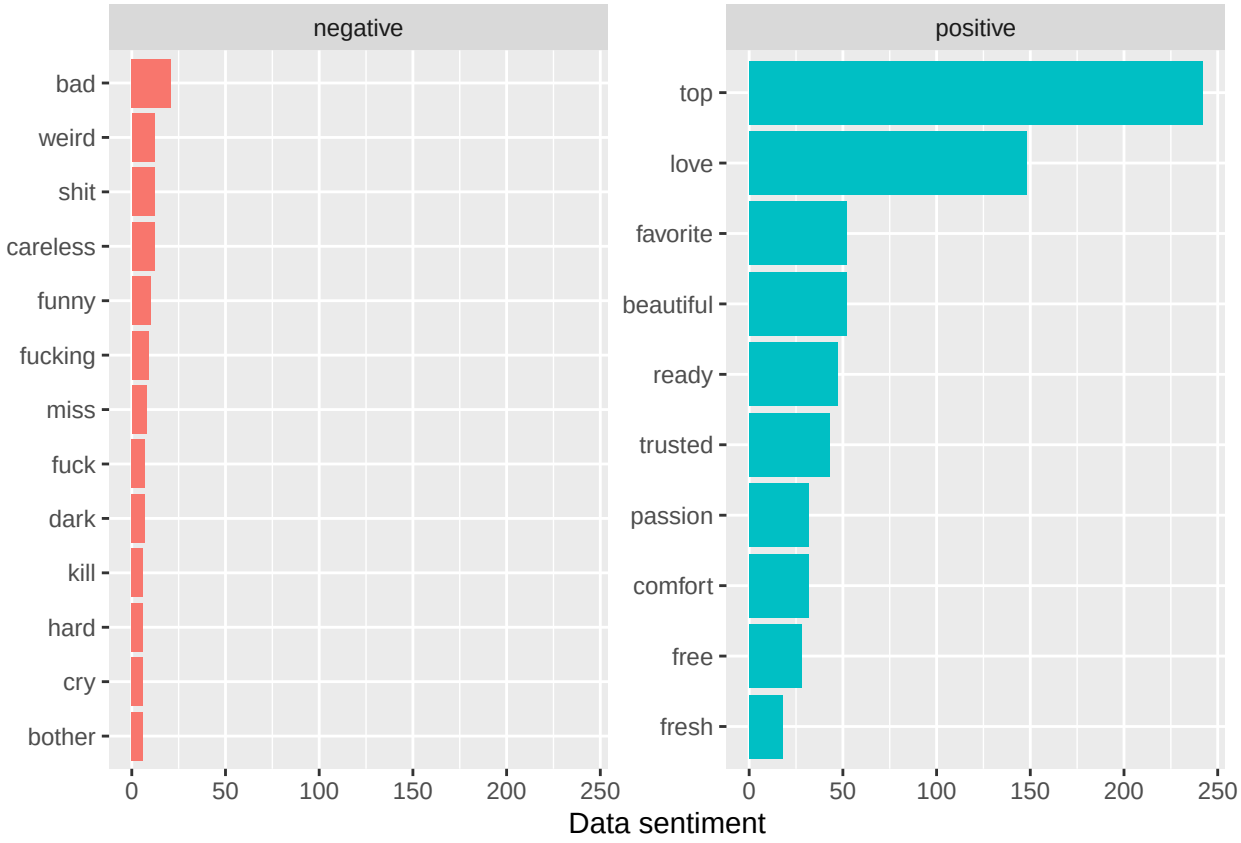


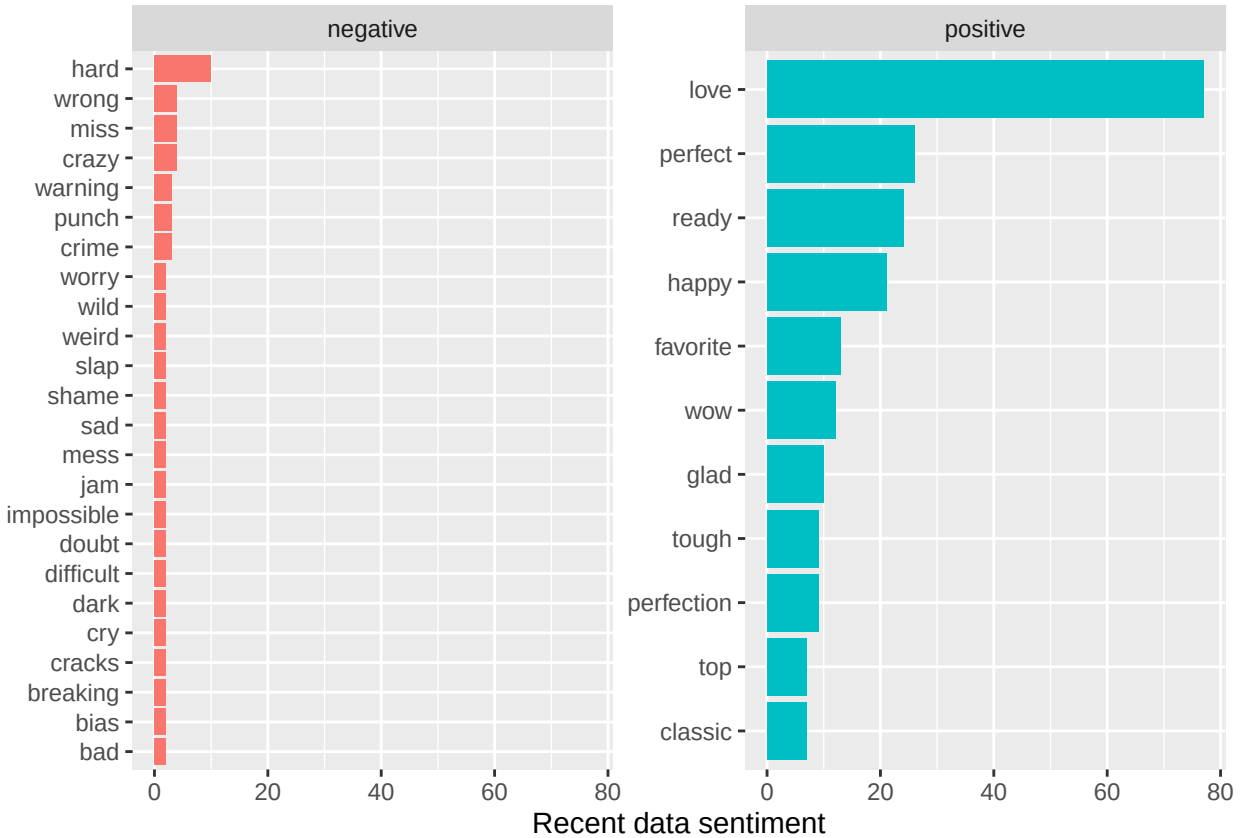
```
## [1] "Mean Original sentiment score = 0.207139265962795"
```

```
## [1] "Mean Working Original sentiment score = 0.601208459214502"
```

For the original twitter data, the average emotional score of recent data is much higher than that of a large number of spotify companies we have chosen.

reply data


negative
positive
bad
weird
shit
careless
funny
fucking
miss
fuck
dark
kill
hard
cry
bother
top
love
favorite
beautiful
ready
trusted
passion
comfort
free
fresh
0
50
100
150
200
250
Data sentiment

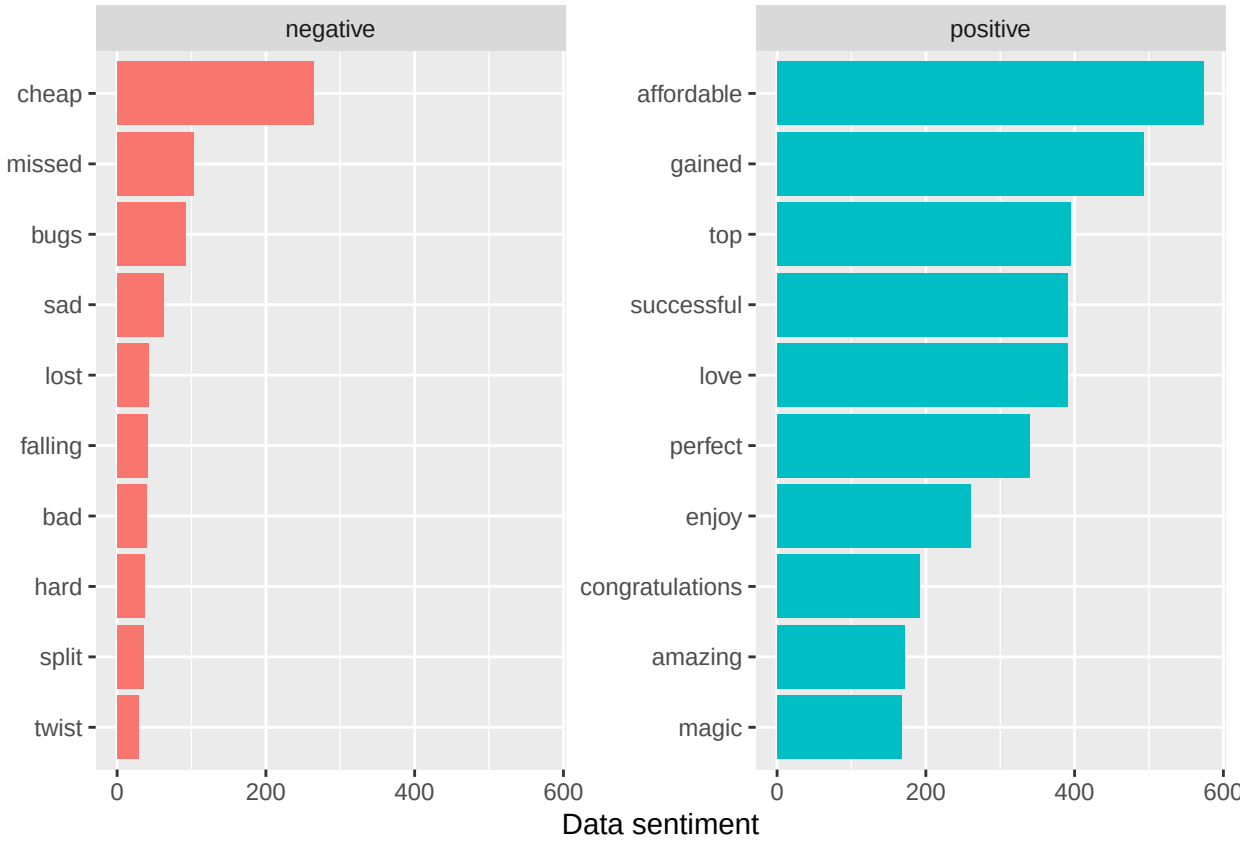

negative
positive
hard
wrong
miss
crazy
warning
punch
crime
worry
wild
weird
slap
shame
sad
mess
jam
impossible
doubt
difficult
dark
cry
cracks
breaking
bias
bad
love
perfect
ready
happy
favorite
wow
glad
tough
perfection
top
classic
0
20
40
60
80
Recent data sentiment

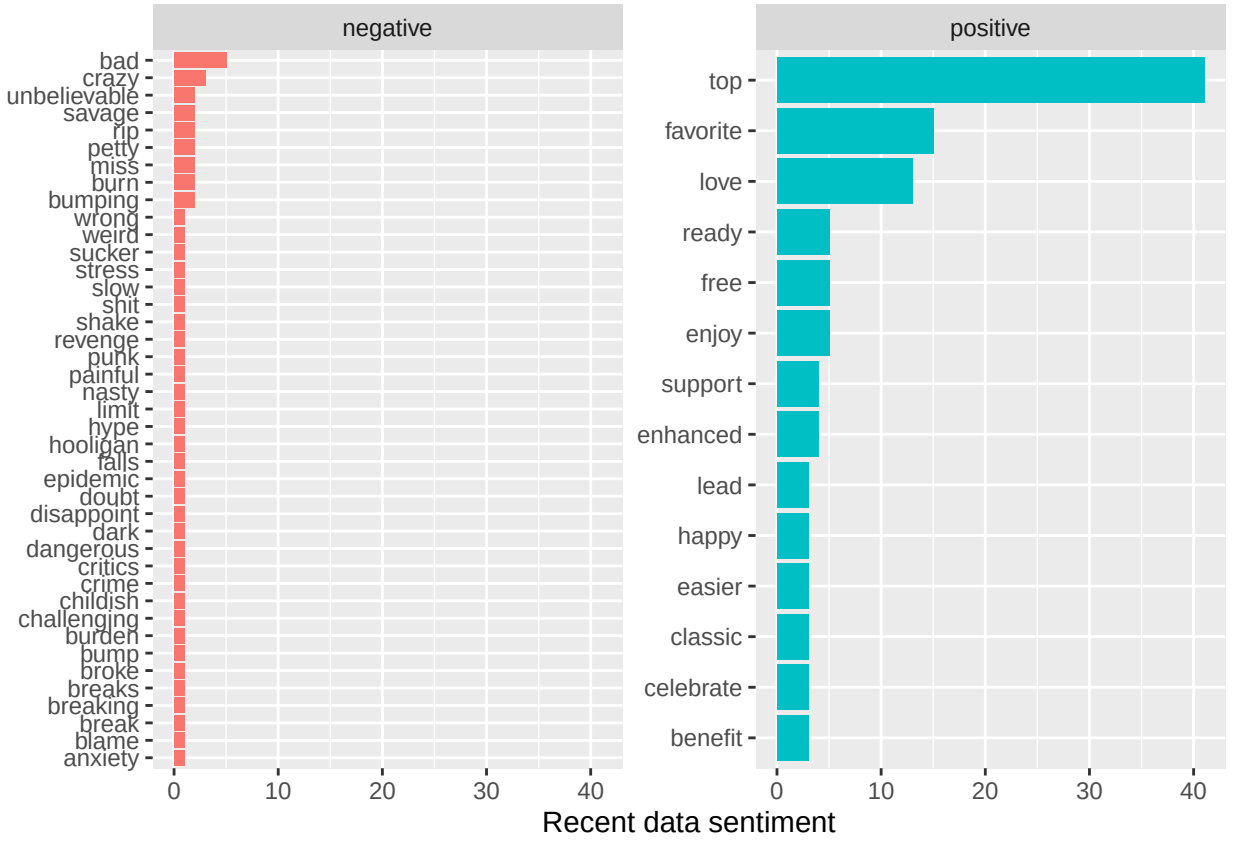
```
## [1] "Mean reply sentiment score = 1.32838589981447"
```

```
## [1] "Mean Working reply sentiment score = 0.628571428571429"
```

However, in terms of twitter response data, the average score of recent data is far lower than that of a large number of spotify companies, which indicates that many negative emotions have been added in recent user responses.

Retweet data

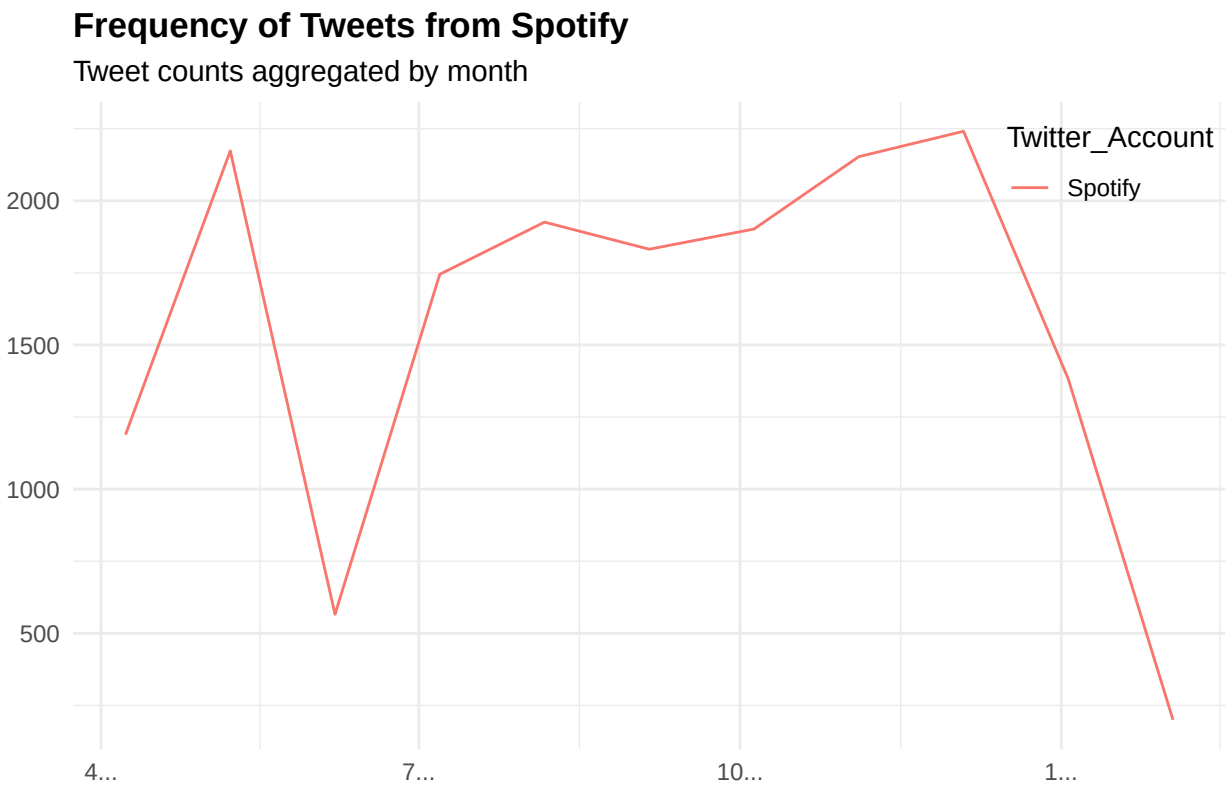




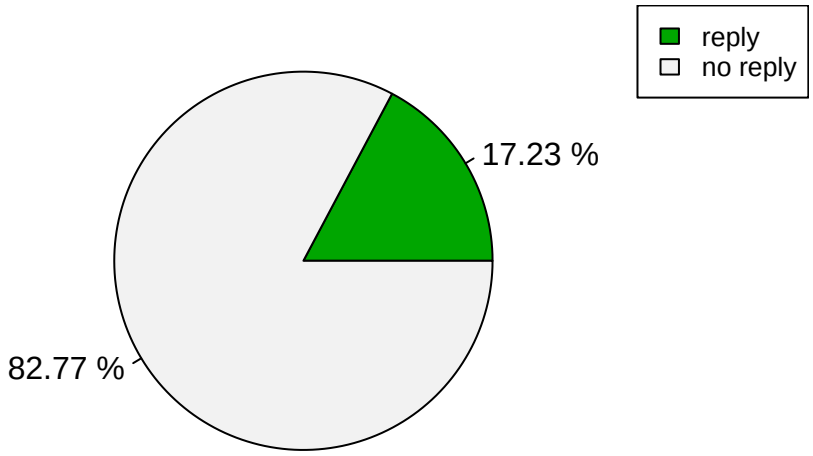
```
## [1] "Mean Retweets sentiment score = 1.17295952973095"
```

```
## [1] "Mean Working Retweets sentiment score = 0.874251497005988"
```

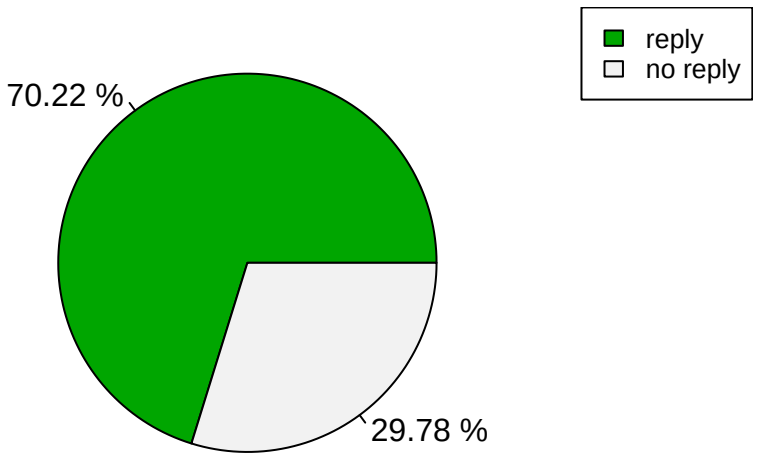
In terms of twitter forwarding data, the average emotional score of recent data is higher than that of a large number of spotify company data we selected.



Source: Data collected from Twitter's REST API via rtweet




```
## [1] "Twitter response rate  = 17.23 %"
```




```
## [1] "Twitter response rate in the past nine months  = 70.22 %"
```

From the comparison of the two pie charts, we can clearly see that the response rate in the last nine months or so has greatly improved compared with the average.