CS 688 Project: Twitter Stock Project Analysis

Emily Darigo
10/14/2019

Stocks Chosen

Gainers:

1. AECOM ($ACM)
2. Reta Pharmaceuticals, Inc. ($RETA)
3. Shopify Inc. ($SHOP)

Losers:

1. Parsley Energy, Inc. ($PE)
2. SmileDirectClub, Inc. ($SDC)
3. CrowdStrike Holdings, Inc. ($CRWD)

- Cashtag was used to pull all tweets
  - Returned more tweets than using company name
- Stocks chosen based on how many tweets were available
  - Some of the top gainers/losers returned less than 10 tweets

Pre-processing Steps

Pre-processing tweets executed in the following order:

1. Remove http elements
2. Transform all words into lowercase
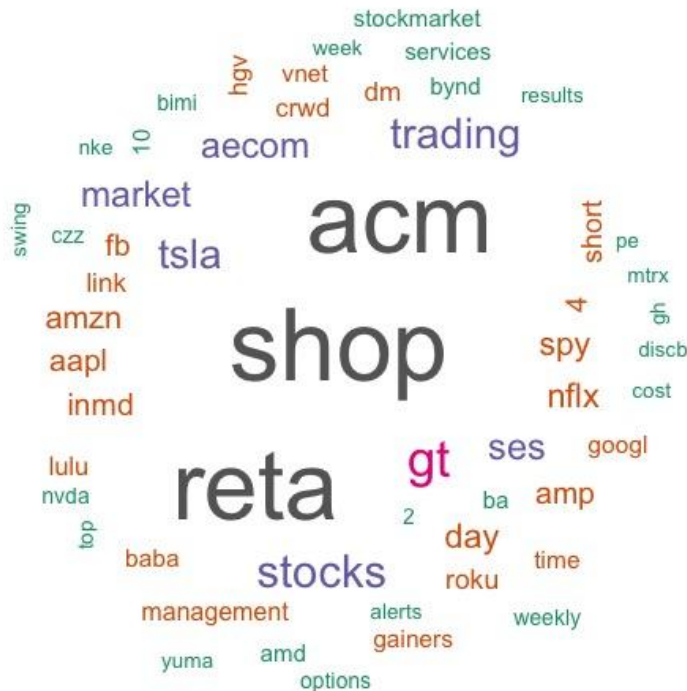3. Remove punctuation
4. Remove English stop-words

```
## Pre-processing Function --------------------------------------
cleaning.function <- function(tweets){
  # Remove http elements manually
  tweets$stripped_text <- gsub("http\\S+","",tweets)
  tweets$stripped_text <- gsub("[^\u0020-\u007F]+","",tweets$stripped_text)
  tweets$stripped_text <- gsub("'I'","",tweets$stripped_text)

  # Lowercase, remove punctuation, remove English stop_words
  data("stop_words")
  tweets.clean <- tweets %>%
    select(stripped_text) %>%
    unnest_tokens(word, stripped_text) %>%
    anti_join(stop_words)
}
```

Most Frequent: Gainers

Top 10 words:

1. shop - 33600
2. acm - 32700
3. reta - 32100
4. gt - 14700
5. stocks - 12600
6. trading - 11400
7. tsla - 10800
8. ses - 9300
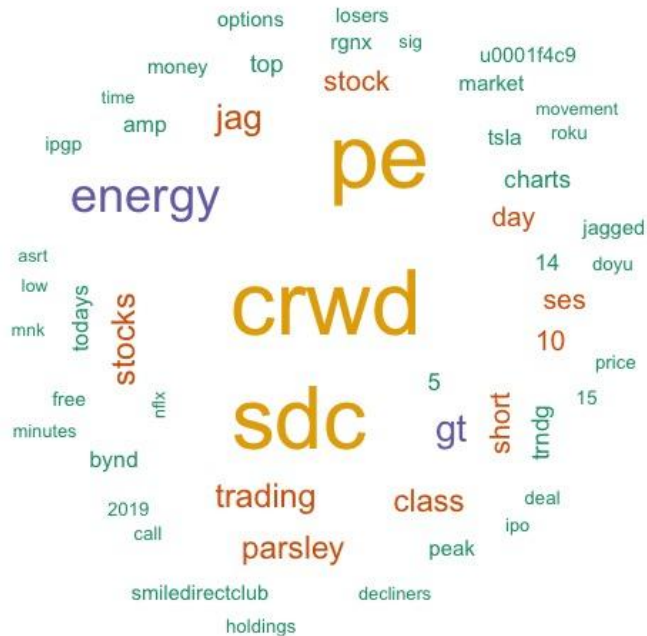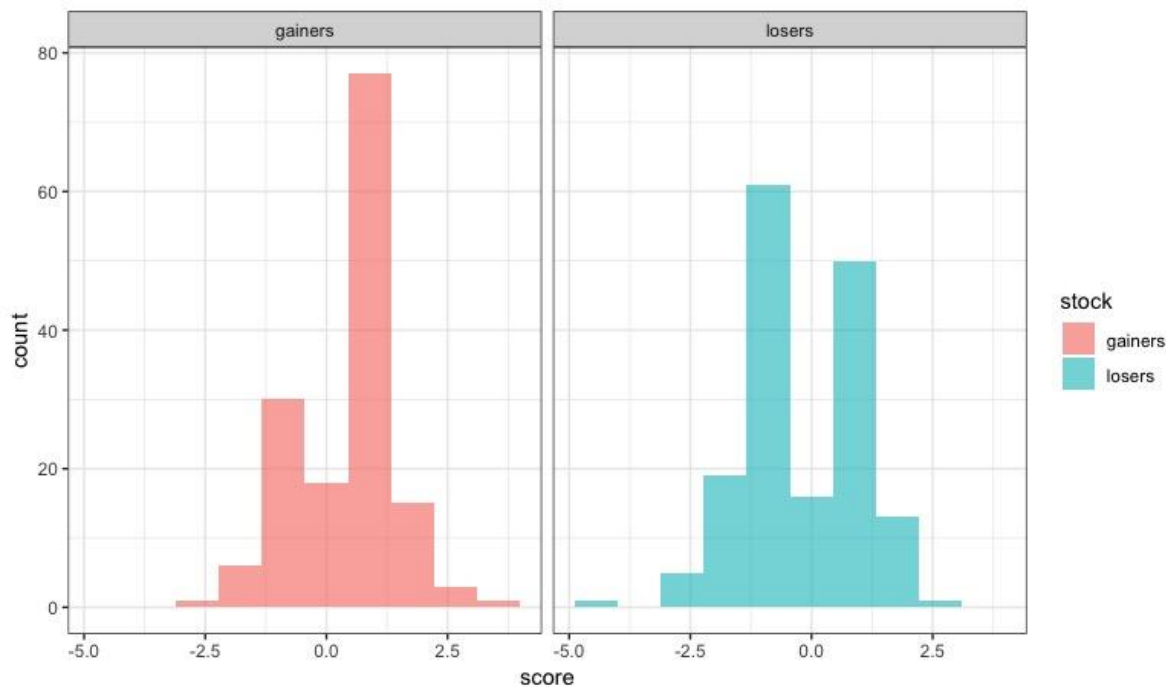9. aecom - 9000
10. market - 9000

Most Frequent: Losers

Top 10 words:

1. crwd - 38400
2. pe - 38400
3. sdc - 36600
4. energy - 18000
5. gt - 12900
6. jag - 11400
7. trading - 10500
8. stocks - 10200
9. parsley - 9900
10. class - 9600

Sentiment Score: Visual comparison

Sentiment Score: Table Summary

Gainers:

- Count: 300
- Median: 1
- SD: 1.17
- Max: 4
- Min: -3

Losers:

- Count: 300
- Median: -1
- SD: 1.35
- Max: 3
- Min: -4

Extra Credit: Data Visualization