

Twitter Text Mining and Stock Price Movement



Jason Zhang
Eric Yang
Chenning Zhang



Research Questions

- 1. Are there correlations between tweets sentiments and stock prices within a week? How are they related?**
- 2. Could tweets sentiments be instructive on predicting stock movement ?**



Data Extraction, Preparation

Acquire Twitter Sentiment

- Collect tweets relevant only to stock ticker
- Twitter API calls through Rtweet package
- Tweets data including user_id, text, symbols included in tweets, and date

Stock Price Price Retrieval

- Collect stock price using Yahoo Finance Data
- Pricing Data including, date, daily high, low, close, trading volume, and adjusted daily price



Data Cleansing

Twitter Data Cleansing

- Remove leading whitespaces
- Remove trailing whitespace
- Remove extra whitespaces
- Replace apostrophes with %'
- Remove emojis
- Remove Unicodes
- Replace double space with single
- Remove URL
- And more

Stock Price Cleansing

- Select closing price and date
- Extract and convert date column

```
x <- tweetsdf
x$text <- enc2native(x$text)
x$text <- gsub("^[:space:]*", "", x$text)
x$text <- gsub("[[:space:]]*$", "", x$text)
x$text <- gsub(" +", " ", x$text)
x$text <- gsub("'", "%'", x$text)
x$text <- iconv(x$text, "latin1", "ASCII", sub="")
x$text <- gsub("<(.*)>", "", x$text)
x$text <- gsub("\\ \\.", " ", x$text)
x$text <- gsub(" ", " ", x$text)
x$text <- gsub("%%", "\\", x$text)
x$text <- gsub("https(.)*$", "", x$text)
x$text <- gsub("\\n", "-", x$text)
x$text <- gsub("--", "-", x$text)
x$text <- gsub("&", "&", x$text)
```



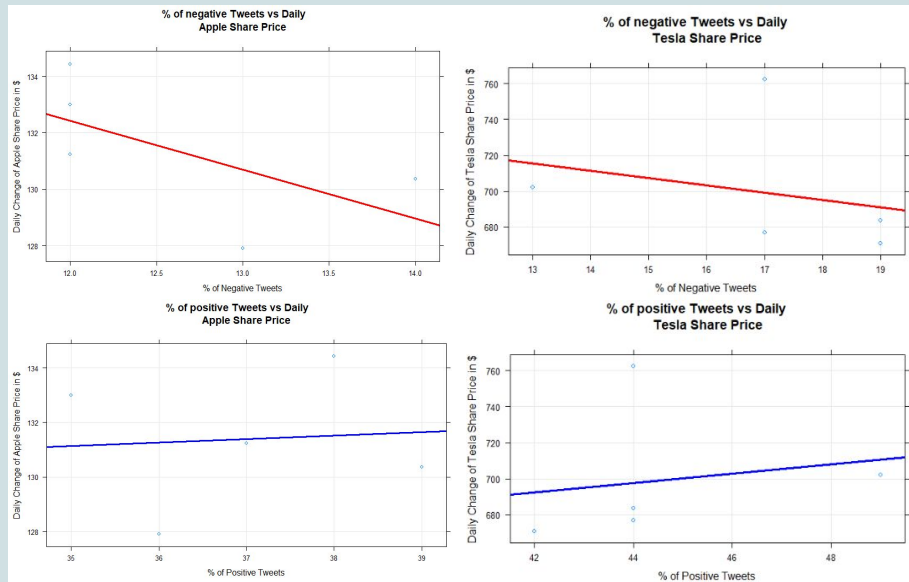
Observation and Conclusion

Correlation Comparison

	Apple	Tesla
Negative Sentiments	-0.68	-0.27
Positive Sentiments	0.08	0.181

Negative sentiments have more predictive power than positive sentiments.

Be aware of the change in public sentiment on social media and adjust your portfolio accordingly.





Recommendation and Limitation

1. Larger datasets are likely to generate deeper level insights
2. Seek for creative ways to quantify sentiments using various dictionaries
3. Include more stock tickers allow to smooth variance in sentiment