

 Overview & Setup Sentiment Analysis

Twitter Sentiment Analysis

Overview & Setup

This notebook walks you through on:

1. How to run a sentiment analysis on tweets using the VADER package
2. How to interpret the VADER scores
3. How to view top positive/negative tweets
4. How to visualize the compound scores distribution

Install and load packages

Install `tidyverse` and `vader` if you do not have them in your R environment.

- `tidyverse` is a collection of R packages for data science
- `vader` (Valence Aware Dictionary and sEntiment Reasoner) is a rule-based sentiment analysis tool specifically attuned to social media text

```
# uncomment and run the lines below if you need to install these packages
# install.packages("tidyverse")
# install.packages("vader")
```

Load packages.

```
library(tidyverse)
library(vader)
```

Read CSV file

```
df_tweets = read_csv('lululemon-tweets.csv')
```

```
## Rows: 4516 Columns: 3
## — Column specification —————
## Delimiter: ","
## chr (2): username, text
## dbl (1): id
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
df_tweets %>% head()
```

	id	username
	<dbl>	<chr>
	1.603435e+18	demsRinsane
	1.603434e+18	sherbertkuma
	1.603433e+18	aidaampie
	1.603433e+18	marinarodich
	1.603433e+18	poshcitystore
	1.603432e+18	marinarodich

6 rows | 1-2 of 3 columns

Print out the number of rows.

```
nrow(df_tweets)
```

```
## [1] 4516
```

Sentiment Analysis

Sample usage

To analyze a piece of text using VADER, use the `get_vader()` function. Here is an example using one of the tweets.

- `compound` score is the “overall” score between -1 (most extreme negative) and +1 (most extreme positive).
- `pos` , `neg` , and `neu` are ratios for proportions of text that fall in each category. These should all add up to be 1.
- We are mainly interested in the compound score.

```
get_vader("Got my BeautyBase merch I am happy I got the first batch 🥰")
```

```
##                               word_scores
## "{0, 0, 0, 0, 0, 0, 2.7, 0, 0, 0, 0, 0, 0}"
##                               compound
##                               "0.572"
##                               pos
##                               "0.236"
##                               neu
##                               "0.764"
##                               neg
##                               "0"
##                               but_count
##                               "0"
```



Calculate scores for all tweets

Because we have a `DataFrame (tibble)` that contains tweet texts in the `text` column, we will need to run `get_vader()` on each row of the `text` column. We can do this by using the `lapply()` function.

The code below stores the VADER results to a list variable named `vscores`. Note that this may take a while (a few minutes) if you have thousands of tweets.

```
vscores <- df_tweets$text %>% lapply(get_vader)
```

```
## Warning in sentiments[i] <- senti_valence(wpe, i, item): number of items to
## replace is not a multiple of replacement length
```

Extract compound score and positive/neutral/negative percentages into separate columns.

```
df_tweets <- df_tweets %>% mutate(
  compound = vscores %>% sapply(function(v) { as.numeric(v["compound"]) }),
  pos = vscores %>% sapply(function(v) { as.numeric(v["pos"]) }),
  neu = vscores %>% sapply(function(v) { as.numeric(v["neu"]) }),
  neg = vscores %>% sapply(function(v) { as.numeric(v["neg"]) }),
)
df_tweets
```

	id	username	
	<dbl>	<chr>	
	1.603435e+18	demsRinsane	
	1.603434e+18	sherbertkuma	
	1.603433e+18	aidaampie	
	1.603433e+18	marinarodich	
	1.603433e+18	poshcitystore	
	1.603432e+18	marinarodich	
	1.603431e+18	Noosiekins	
	1.603431e+18	TroiBear28	
	1.603431e+18	JoeShmo32152716	
	1.603430e+18	ClothingShoesA1	
1-10 of 4,516 rows 1-2 of 7 columns			
		Previous	1 2 3 4 5 6 ... 452 Next



50 most positive tweets

```
df_tweets %>%
  arrange(desc(compound)) %>%
  select(text, username, compound, pos, neu, neg) %>%
  head(50)
```

text

<chr>

@CrystalRunsFree @micmacmall @lululemon @ROOTS Super fun, Crystal! Never too old (age is just a play/be playful and just have fun doing whatever activities you enjoy!😊 Happy to see that you were o getting your holiday cheers on! 🥰

🔴 LULULEMON - NOW OPEN! 🔴 \n\n@lululemon is here, and we're living for it! Come check out the selection of high-quality athletic wear and gear. Their knowledgeable and friendly staff are on hand to your perfect outfit or a special holiday gift! <https://t.co/WFyNjHwJ5E>

@desota @lululemon @Costco @Broadcom is taking over the blockchain, \nA cool token with 0% tax, presale is live, with an active dev, best community ever, we moon with winners, join the winning team #Eth tested and trusted✅✅ to the moon\nn<https://t.co/Onkft3RvZc>\n<https://t.co/1UkcFgSko4>

@eWhispers @autozone @Veru_Pharma @lululemon @DocuSign @MongoDB @SAICinc @DLHCorp @ @SentinelOne @gitlab is taking over the blockchain, \nA cool token with 0% tax, 100%, presale is live active dev, best community ever, we moon with winners, join the winning team today #bnb #Eth tested trusted✅✅ to the moon\nn<https://t.co/Onkft3RvZc>\n<https://t.co/1UkcFgSko4>

@desota @netflix @lululemon @DocuSign is taking over the blockchain, \nA cool token with 0% tax, 1 presale is live, with an active dev, best community ever, we moon with winners, join the winning team #Eth tested and trusted✅✅\n<https://t.co/Onkft3AsXc>\nthe moon is our destination ✅🚀🚀
<https://t.co/Dzat8u34eN>

@calvinmcdonald @lululemon Hello Mr Mcdonald. Love your products! Wow! Question. My wife looks the leggings you make... i figure you owe me a commission on all the sales she causes. Lol. Merry Ch

@SavLovesSwift I love aerie leggings mine are SO soft, but honestly lululemon, they are expensive bu everything I've gotten from there has been honestly worth the money IMO

@Adriana1o5 We do that, same with PPC. We run ab 2M a year through them but what I am talking at the random benefits like free uber eats credits, 10% off lululemon, etc... all these benefits are on thing personally, not for biz.

@Amwysoc It is a great mall . My fav too ... live 15 minutes from it . But 2 joggers at Lululemon ... \$27! year old is lucky I love him ! Ha !

OMG I love LOVE Lululemon! Please don't tell me it's going out of business. Ladies, they sell thee bes <https://t.co/ctvQsq43wr>

1-10 of 50 rows | 1-1 of 6 columns

Previous 1 2 3 4 5 Next

👎 50 most negative tweets

```
df_tweets %>%
  arrange(compound) %>%
  select(text, username, compound, pos, neu, neg) %>%
  head(50)
```

text

<chr>

You are a thief, Ronna! You are stealing American's money to buy Lululemon? In case no one has told way too fat to wear that brand. Again, stealing and wasting our money. RESIGN. <https://t.co/pCulKm6>

text

<chr>

Just had the worst experience with @lululemon Such a big company but totally ignores their customer no responsibility for their errors. So disappointed 😞 #lululemon #lululemoncanada

This crime is insane and terrifying... tune in to hear Destiney delve into the Lululemon murder. \n\nSpoc
\n\nhttps://t.co/nnDTfZtb32\n\nAmazon Music:\n\nhttps://t.co/FVHI2w8PBO\n\nApple
Podcast:\n\nhttps://t.co/hqRRPRZgeg https://t.co/jKPWkrcWtK

@lululemon and @LaserShip shipments to Miami are definitely being stolen. Check the google reviews delivered but not actually delivered. SEVERELY disappointed. I had a package worth over \$200 stolen #Stolen #Lululemon

@TuckerPeterson_ @KurtSchlichter @robbystarbucks @GOPChairwoman Aw, hell no. If I go so far as to RNC versus an individual candidate, and I have in the past, I'll be damned pissed to find out it got spe freaking Lululemon.

@SquawkStreet @lululemon @calvinmcdonald @jimcramer U.S. to send \$275 million in military aid to NATO chief worries war could become wider conflict\n\nBlood on your hands u war mongers

@rgkicks @HButter03 @ApeCrayonsYummy @unusual_whales Stop the cap with your faded ass under just because you're too poor to afford lululemon

@nationalpost @lululemon stuff is now sweat shop fast fashion crap AND they are appealing to a woke ALL of their marketing rather than their actual consumer base. Fools fools fools

It's hard shopping from lululemon without thinking about the lululemon murders \U{01fae3} gotta stop crime shit frfr

Does Lululemon's 12% Drop Signal Bad News For Clothing Retailers?\n\nLuxury athletic clothing retailer Lululemon gapped down 12%, after warning of higher-than-anticipated inventories. Is this a bad sign for clothing stocks? https://t.co/4ksRMMZxYI

1-10 of 50 rows | 1-1 of 6 columns

Previous 1 2 3 4 5 Next

**Average compound score**

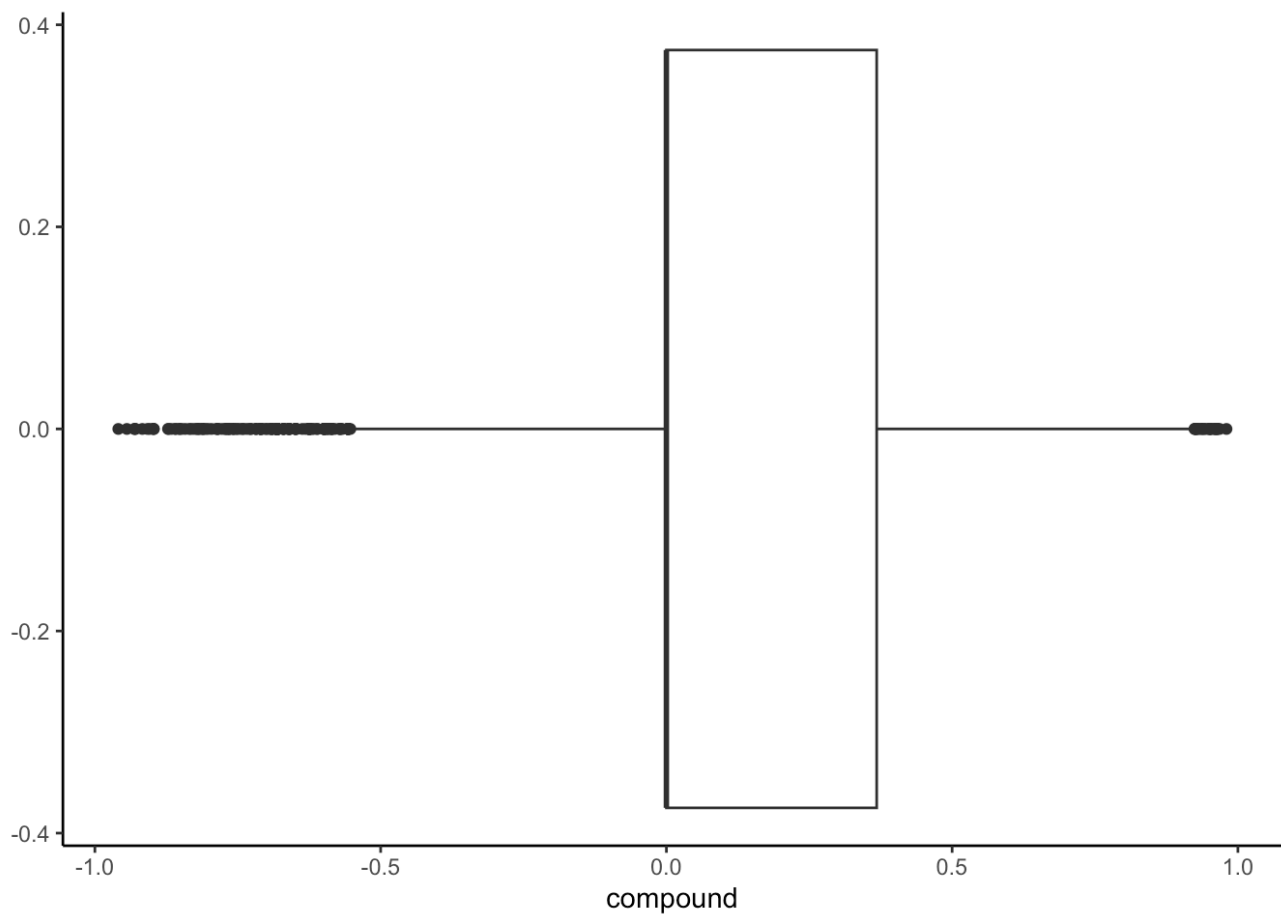
```
mean(df_tweets$compound)
```

```
## [1] 0.1257358
```

**Box plot of compound scores**

The box plot below gives you the spread and skewness of compound scores.

```
df_tweets %>%
  ggplot(aes(x=compound)) +
  theme_classic() +
  geom_boxplot()
```



Histogram of compound scores

The histogram below gives you an approximate distribution of the compound scores.

```
df_tweets %>%  
  ggplot(aes(x=compound)) +  
  theme_classic() +  
  geom_histogram(bins = 20, color="black", fill="white")
```

