TwitterAnalysis

nsardar

2022-07-23

Loading the libraries

```
options(scipen = 999) #avoid the scientific notation
library(twitteR)
library(writexl)
library(bitops)
library(tidytext)
library(devtools)
library(sentiment)
library(quanteda)
library(quanteda.textstats)
library(quanteda.textplots)
```

Getting Data

Based on the tweets from 4 different mobile manufacturing companies and then we will be doing text analysis and visualization for this companies.

We will be using twitter developer portal to get acces to tweets from twitter website.

```
consumer_key = "3N1fyHxwhVAVs5BCXzuQ8vbXV"
consumer_secret = "IVf6acHFxRWVQ006HNlDP9hygzOXx1XkZdKF9cMZlbL717AIQu"
access_token = "1458218683017347078-UTwl4ULAneM8wgDJypPXsU2KNhI8pU"
acess_token_secret = "4mLTYRLgJHSEoeV3qdz10WT7NdjwOH95AgvA8QCULhr0e"

setup_twitter_oauth(consumer_key,consumer_secret,access_token,acess_token_secret)
```

```
## [1] "Using direct authentication"
```

Mobile Companies

We will be using data from the below mentioned companies : 1) Motorola 2) Xiaomi 3) Oneplus 4) Samsung

```
tweetsMotorola = userTimeline('MotorolaUS', n = 1000)
tweetsXiaomi = userTimeline('Xiaomi', n = 1000)
tweetsOnePlus = userTimeline('oneplus', n = 1000)
tweetsSamsung = userTimeline('SamsungMobile', n = 1000)


tweetsDfMotorola = twListToDF(tweetsMotorola)
head(tweetsDfMotorola,2)
```

```
##
text
## 1 @JoeStep92374964 So sorry for the inconvenience this may have caused. Since you have mentioned that you have spoken… ht
tps://t.co/R17f6Bjvlq
## 2                                @JPrice02 We're sorry for any inconvenience. If you wish to reconsider, we are standi
ng by to help. -Erwin
##    favorited favoriteCount        replyToSN           created truncated
## 1     FALSE             0 JoeStep92374964 2022-07-23 13:06:04      TRUE
## 2     FALSE             0         JPrice02 2022-07-23 00:51:55     FALSE
##         replyToSID               id replyToUID
## 1 1550718119001427968 1550829601890467841 1594103780
## 2 1550642166065659904 1550644844086820864   22788127
##                                                 statusSource screenName
## 1 <a href="https://www.sprinklr.com" rel="nofollow">Sprinklr</a> MotorolaUS
## 2 <a href="https://www.sprinklr.com" rel="nofollow">Sprinklr</a> MotorolaUS
##    retweetCount isRetweet retweeted longitude latitude
## 1            0     FALSE     FALSE        NA       NA
## 2            0     FALSE     FALSE        NA       NA
```

```
tweetsDfXiaomi = twListToDF(tweetsXiaomi)
head(tweetsDfXiaomi,2)
```

```
##
text
## 1 It is a beautiful thing to discover your own approach to style and hear your soul speak. #Xiaomi12Lite… https://t.co/L3
1RfFZybZ
## 2        Look exquisite, be confident, and stay determined with #Xiaomi12Lite at your side. #MyStyleMyShot https://t.co/uT
VwCxkgtW
##    favorited favoriteCount replyToSN           created truncated replyToSID
## 1     FALSE           245      <NA> 2022-07-23 13:00:12     TRUE       <NA>
## 2     FALSE           431      <NA> 2022-07-22 13:00:10    FALSE       <NA>
##                   id replyToUID
## 1 1550828122404167680       <NA>
## 2 1550465728775704584       <NA>
##                                                                  statusSource
## 1 <a href="https://prod1.sprinklr.com" rel="nofollow">Sprinklr Publishing</a>
## 2 <a href="https://prod1.sprinklr.com" rel="nofollow">Sprinklr Publishing</a>
##    screenName retweetCount isRetweet retweeted longitude latitude
## 1     Xiaomi           25     FALSE     FALSE        NA       NA
## 2     Xiaomi           38     FALSE     FALSE        NA       NA
```

```
tweetsDfOnePlus = twListToDF(tweetsOnePlus)
head(tweetsDfOnePlus,2)
```

```
##                                                                          text
## 1               Low light = blurry nights? Keep things sharp with the #OnePlus10T <U+0001F4F8>
## 2 What went into the #OnePlus10T's camera? Let's find out<U+0001F399><U+FE0F> https://t.co/8AJ7AF3Ojd
##    favorited favoriteCount replyToSN           created truncated replyToSID
## 1     FALSE           176      <NA> 2022-07-23 14:00:01    FALSE       <NA>
## 2     FALSE           182      <NA> 2022-07-23 11:00:00    FALSE       <NA>
##                   id replyToUID
## 1 1550843179389227010       <NA>
## 2 1550797874933276677       <NA>
##                                                                  statusSource
## 1 <a href="https://twitter.com" rel="nofollow">Twitter for Advertisers</a>
## 2  <a href="https://mobile.twitter.com" rel="nofollow">Twitter Web App</a>
##    screenName retweetCount isRetweet retweeted longitude latitude
## 1     oneplus           13     FALSE     FALSE        NA       NA
## 2     oneplus           16     FALSE     FALSE        NA       NA
```

```
tweetsDfSamsung = twListToDF(tweetsSamsung)
head(tweetsDfSamsung,2)
```

```
##
text
## 1         Privacy is nothing without strong security. Your privacy. Secured. #SamsungKnox\n\nWhich feature makes you
feel safe on your phone?
## 2 "What was once a novelty three years ago, is now the preferred choice for millions." TM Roh, President &amp; Head of M
X… https://t.co/9CJngoKDK0
##    favorited favoriteCount replyToSN           created truncated replyToSID
## 1     FALSE           206      <NA> 2022-07-22 10:00:01    FALSE       <NA>
## 2     FALSE           292      <NA> 2022-07-21 01:07:52     TRUE       <NA>
##                   id replyToUID
## 1 1550420391751045122       <NA>
## 2 1549924084913688576       <NA>
##                                                                  statusSource
## 1         <a href="https://bn.co" rel="nofollow">Brand Networks Composer</a>
## 2 <a href="https://studio.twitter.com" rel="nofollow">Twitter Media Studio</a>
##      screenName retweetCount isRetweet retweeted longitude latitude
## 1 SamsungMobile           24     FALSE     FALSE        NA       NA
## 2 SamsungMobile           31     FALSE     FALSE        NA       NA
```

Sentiment Analysis

The below score indicates whether the tweets are positive , negative or neutral

4 Positive

2 Neutral

0 Negative

```
finalExtractedTweets = rbind(tweetsDfMotorola, tweetsDfXiaomi, tweetsDfOnePlus, tweetsDfSamsung)

encodeSource <- function(x) {
  if(x=="<a href=\"http://twitter.com/download/iphone\" rel=\"nofollow\">Twitter for iPhone</a>"){
    gsub("<a href=\"http://twitter.com/download/iphone\" rel=\"nofollow\">Twitter for iPhone</a>", "iphone", x,fixed=TRUE)
  }else if(x=="<a href=\"http://twitter.com/#!/download/ipad\" rel=\"nofollow\">Twitter for iPad</a>"){
    gsub("<a href=\"http://twitter.com/#!/download/ipad\" rel=\"nofollow\">Twitter for iPad</a>","ipad",x,fixed=TRUE)
  }else if(x=="<a href=\"http://twitter.com/download/android\" rel=\"nofollow\">Twitter for Android</a>"){
    gsub("<a href=\"http://twitter.com/download/android\" rel=\"nofollow\">Twitter for Android</a>","android",x,fixed=TRUE)
  } else if(x=="<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client</a>"){
    gsub("<a href=\"http://twitter.com\" rel=\"nofollow\">Twitter Web Client</a>","Web",x,fixed=TRUE)
  } else if(x=="<a href=\"http://www.twitter.com\" rel=\"nofollow\">Twitter for Windows Phone</a>"){
    gsub("<a href=\"http://www.twitter.com\" rel=\"nofollow\">Twitter for Windows Phone</a>","windows phone",x,fixed=TRUE)
  }else if(x=="<a href=\"http://dlvr.it\" rel=\"nofollow\">dlvr.it</a>"){
    gsub("<a href=\"http://dlvr.it\" rel=\"nofollow\">dlvr.it</a>","dlvr.it",x,fixed=TRUE)
  }else if(x=="<a href=\"http://ifttt.com\" rel=\"nofollow\">IFTTT</a>"){
    gsub("<a href=\"http://ifttt.com\" rel=\"nofollow\">IFTTT</a>","ifttt",x,fixed=TRUE)
  }else if(x=="<a href=\"http://earthquaketrack.com\" rel=\"nofollow\">EarthquakeTrack.com</a>"){
    gsub("<a href=\"http://earthquaketrack.com\" rel=\"nofollow\">EarthquakeTrack.com</a>","earthquaketrack",x,fixed=TRUE)
  }else if(x=="<a href=\"http://www.didyoufeel.it/\" rel=\"nofollow\">Did You Feel It</a>"){
    gsub("<a href=\"http://www.didyoufeel.it/\" rel=\"nofollow\">Did You Feel It</a>","did_you_feel_it",x,fixed=TRUE)
  }else if(x=="<a href=\"http://www.mobeezio.com/apps/earthquake\" rel=\"nofollow\">Earthquake Mobile</a>"){
    gsub("<a href=\"http://www.mobeezio.com/apps/earthquake\" rel=\"nofollow\">Earthquake Mobile</a>","earthquake_mobile",x,
fixed=TRUE)
  }else if(x=="<a href=\"http://www.facebook.com/twitter\" rel=\"nofollow\">Facebook</a>"){
    gsub("<a href=\"http://www.facebook.com/twitter\" rel=\"nofollow\">Facebook</a>","facebook",x,fixed=TRUE)
  }else {
    "others"
  }
}

finalExtractedTweets$tweetSource = sapply(finalExtractedTweets$statusSource,function(sourceSystem) encodeSource(sourceSyste
m))

tweetsSentiment<-sentiment(finalExtractedTweets$text)    # instead used amazondf$text
finalExtractedTweets$polarity <- tweetsSentiment$polarity # Polarity denotes whther the tweet is neagtive, positive or neutr
al
```

Final Results for Sentiment analysis on tweets

```
head(finalExtractedTweets,5)
```

```
## 
text
## 1 @JoeStep92374964 So sorry for the inconvenience this may have caused. Since you have mentioned that you have spoken… ht
tps://t.co/R17f6Bjvlq
## 2                              @JPrice02 We're sorry for any inconvenience. If you wish to reconsider, we are standi
ng by to help. -Erwin
## 3                                    @JPrice02 Please let us know if you need help arranging th
e repair (2/2) -Erwin
## 4 @JPrice02 We understand. A swollen battery is not a safety issue. It is rare but typically occurs on older batterie… ht
tps://t.co/mUXwIF3DvD
## 5                  @JPrice02 Hi, we're sorry to hear that. May ask if you are still having issues with your apps and ga
metime? Thanks -Erwin
##   favorited favoriteCount        replyToSN              created truncated
## 1     FALSE             0 JoeStep92374964 2022-07-23 13:06:04      TRUE
## 2     FALSE             0         JPrice02 2022-07-23 00:51:55     FALSE
## 3     FALSE             0         JPrice02 2022-07-23 00:35:35     FALSE
## 4     FALSE             0         JPrice02 2022-07-23 00:35:07      TRUE
## 5     FALSE             0         JPrice02 2022-07-22 23:44:56     FALSE
##          replyToSID                  id replyToUID
## 1 1550718119001427968 1550829601890467841 1594103780
## 2 1550642166065659904 1550644844086820864   22788127
## 3 1550633911218323457 1550640736571359233   22788127
## 4 1550633911218323457 1550640617901940736   22788127
## 5 1550620425583038464 1550627990064828418   22788127
##                                                 statusSource screenName
## 1 <a href="https://www.sprinklr.com" rel="nofollow">Sprinklr</a> MotorolaUS
## 2 <a href="https://www.sprinklr.com" rel="nofollow">Sprinklr</a> MotorolaUS
## 3 <a href="https://www.sprinklr.com" rel="nofollow">Sprinklr</a> MotorolaUS
## 4 <a href="https://www.sprinklr.com" rel="nofollow">Sprinklr</a> MotorolaUS
## 5 <a href="https://www.sprinklr.com" rel="nofollow">Sprinklr</a> MotorolaUS
##   retweetCount isRetweet retweeted longitude latitude tweetSource polarity
## 1            0     FALSE     FALSE        NA       NA      others  neutral
## 2            0     FALSE     FALSE        NA       NA      others  neutral
## 3            0     FALSE     FALSE        NA       NA      others  neutral
## 4            0     FALSE     FALSE        NA       NA      others negative
## 5            0     FALSE     FALSE        NA       NA      others positive
```

Extracting Results into excel to visualize in Tableau

```
#write_xlsx(finalExtractedTweets,'TweetsData.xlsx')
```

Text Analysis and Visualization

We will be first visualizing text for individual companies and generate some isnights

```
ourCorpusMotorola =  corpus(tweetsDfMotorola,text_field = 'text')
ourCorpusXiaomi = corpus(tweetsDfXiaomi,text_field = 'text')
ourCorpusOnePlus =  corpus(tweetsDfOnePlus,text_field = 'text')
ourCorpusSamsung =  corpus(tweetsDfSamsung,text_field = 'text')

# Lets remove few stop words and punctuation

ourTokensMotorola <- tokens(ourCorpusMotorola, remove_punct = TRUE)
ourTokensXiaomi <- tokens(ourCorpusXiaomi, remove_punct = TRUE)
ourTokensOnePlus <- tokens(ourCorpusOnePlus, remove_punct = TRUE)
ourTokensSamsung <- tokens(ourCorpusSamsung, remove_punct = TRUE)

ourTokensMotorolaNS <- tokens_select(ourTokensMotorola, stopwords('en'),selection = 'remove' , min_nchar =4)
ourTokensXiaomiNS <- tokens_select(ourTokensXiaomi, stopwords('en'),selection = 'remove' , min_nchar =4)
ourTokensOnePlusNS <- tokens_select(ourTokensOnePlus, stopwords('en'),selection = 'remove' , min_nchar =4)
ourTokensSamsungNS <- tokens_select(ourTokensSamsung, stopwords('en'),selection = 'remove' , min_nchar =4)

ourMatrixMotorola <- dfm(ourTokensMotorolaNS)
ourMatrixXiaomi <- dfm(ourTokensXiaomiNS)
ourMatrixOnePlus <- dfm(ourTokensOnePlusNS)
ourMatrixSamsung <- dfm(ourTokensSamsungNS)

# Top 5 features for each company
topfeatures(ourMatrixMotorola,5)   # Motorola
```

```
##    thanks    sorry reaching    phone     help
##       229      168      144      109      101
```

```
topfeatures(ourMatrixXiaomi,5)     # Xiaomi
```

```
##             xiaomi #risetothechallenge   #mastereveryscene              check
##                121                  81                  56                 52
##  #redminote11series
##                 50
```

```
topfeatures(ourMatrixOnePlus,5)    # Oneplus
```

```
## #oneplus10pro        oneplus    @oneplus_in #shotononeplus           find
##            61             45             32             30             16
```

```
topfeatures(ourMatrixSamsung,5)    # Samsung
```

```
##      #galaxys22 #samsungunpacked           ultra          series
##            289             245             159             139
##        @bts_twt
##             99
```

```
# Wordcloud for most used words in the 5 companies

textplot_wordcloud(ourMatrixMotorola,max_words = 200)  # Motorola
```

```
textplot_wordcloud(ourMatrixXiaomi,max_words = 200)    # Xiaomi
```



```
textplot_wordcloud(ourMatrixOnePlus,max_words = 200)    # Oneplus
```

```
textplot_wordcloud(ourMatrixSamsung,max_words = 200)        # Samsung
```


Let's visualize plot for each company based on frequency of words which has been repeated atleast 10 times.

```
# Motorola
dfmIaugMoto = dfm(ourMatrixMotorola)
dfmIaugMoto = dfm_trim(dfmIaugMoto , min_termfreq=10)
textplot_wordcloud(dfmIaugMoto,max_words = 200 , min_size = 2)
```

```
# Xiaomi
dfmIaugXiaomi = dfm(ourMatrixXiaomi)
dfmIaugXiaomi = dfm_trim(dfmIaugXiaomi , min_termfreq=10)
textplot_wordcloud(dfmIaugXiaomi,max_words = 200 , min_size = 2)
```



```
# Oneplus
dfmIaugOnePlus = dfm(ourMatrixOnePlus)
dfmIaugOnePlus = dfm_trim(dfmIaugOnePlus , min_termfreq=10)
textplot_wordcloud(dfmIaugOnePlus,max_words = 200 , min_size = 2)
```

```
# Samsung
dfmIaugSamsung = dfm(ourMatrixSamsung)
dfmIaugSamsung = dfm_trim(dfmIaugSamsung , min_termfreq=10)
textplot_wordcloud(dfmIaugSamsung,max_words = 200 , min_size = 2)
```


Visualization of all the 5 twitter accounts together

```
our_corpus <- corpus(finalExtractedTweets,text_field = 'text')
head(our_corpus,2)
```

```
## Corpus consisting of 2 documents and 17 docvars.
## text1 :
## "@JoeStep92374964 So sorry for the inconvenience this may hav..."
##
## text2 :
## "@JPrice02 We're sorry for any inconvenience. If you wish to ..."
```

```
# Lets remove few words and punctuation

our_tokens <- tokens(our_corpus, remove_punct = TRUE)
words_to_be_removed<-c() # Currently we are not removing any additional extra words
our_tokens_nonstopwords<- tokens_select(our_tokens, pattern = c(stopwords('en'), words_to_be_removed),selection = 'remove',
 min_nchar = 4)



# Document feature Matrix to show frequency of tokens in each document

our_matrix<-dfm(our_tokens_nonstopwords)
our_matrix<-dfm_trim(our_matrix, min_termfreq = 10, min_docfreq = 5)
nfeat(our_matrix) #number of tokens in all documents
```

```
## [1] 475
```

```
#Most mentioned words

topfeatures(our_matrix,10)
```

```
##          #galaxys22          thanks #samsungunpacked           sorry
##                 289             274             245             169
##               ultra            like          series        reaching
##                 162             154             151             150
##               phone            help
##                 150             137
```

```
topfeatures(our_matrix,30)
```

```
##          #galaxys22          thanks    #samsungunpacked            sorry
##                 289             274                 245              169
##               ultra            like              series         reaching
##                 162             154                 151              150
##               phone            help                know           xiaomi
##                 150             137                 129              121
##              please            hear               check            learn
##                 117             117                 112              108
##            @bts_twt          galaxy          understand            share
##                  99              91                  89               87
##               thank            time #risetothechallenge       experience
##                  85              84                  81               78
##                look            best               hello            ready
##                  75              75                  74               74
##                epic            need
##                  73              71
```

```
set.seed(42)
textplot_wordcloud(our_matrix, max_words = 200, min_size = 1, max_size = 4)  # All 5 companies combined
```

Better Visualization of wordcloud

for all the 5 companies

```
# Text plot showing visualization of all the companies together

corpus_subset(our_corpus, screenName %in% c("MotorolaUS", "Xiaomi", "oneplus","SamsungMobile")) %>%
  tokens(remove_punct = TRUE) %>%
  tokens_remove(stopwords("english")) %>%
  dfm() %>%
  dfm_group(groups = screenName) %>%
  dfm_trim(min_termfreq = 5, verbose = FALSE) %>%
  textplot_wordcloud(comparison = TRUE, max_words = 200)
```