Projeto - Data Science Academy

André Campos da Silva

10 de Novembro, 2020

Projeto - Analise de Sentimentos

Fazer uma analise de sentimentos no Twitter sobre as eleições dos USA.

Coletando os dados

```
# Carrego os pacotes necessários para o projeto
#install.packages("twitteR")
#install.packages("ROAuth")
#install.packages('tm')
#install.packages("syuzhet")
#install.packages("wordcloud")

library('wordcloud')
```

```
## Loading required package: RColorBrewer
```

```
library("twitteR")
library("ROAuth")
library("tm")
```

```
## Loading required package: NLP
```

```
library("syuzhet")
library('tidyverse')
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x ggplot2::annotate() masks NLP::annotate()
## x dplyr::filter()     masks stats::filter()
## x dplyr::id()         masks twitteR::id()
## x dplyr::lag()        masks stats::lag()
## x dplyr::location()   masks twitteR::location()
```

```
# Criar as chaves e faço a conexão para o Twitter
api_key <- 'xxxxxxxxxxxxxxxxxxxxxxx'
api_key_secret <- 'xxxxxxxxxxxxxxxxxxxxxx'
access_token <-'xxxxxxxxxxxxxxxxxxxxxx-xxxxxxxxxxxxxxxxxxxxx'
access_token_secret <-'xxxxxxxxxxxxxxxxxxxxxx'
```

```
setup_twitter_oauth(consumer_key = api_key, consumer_secret = api_key_secret,
                    access_token = access_token, access_secret = access_token_secret)
```

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

Tratamento dos dados

```
# Extraio 500 twitters em ingles com a #Joe Biden a intenção de verificar o que as pessoas
# estão falando a respeito dele e da sua vitoria na eleição.

twtters <- searchTwitter('#Joe Biden', n = 500, lang = 'en')
```

```
## Warning in doRppAPICall("search/tweets", n, params = params, retryOnRateLimit =
## retryOnRateLimit, : 500 tweets were requested but the API can only return 295
```

```
electionUSA <- twListToDF(twtters)
head(electionUSA)
```

```
##
text
## 1       RT @BorjeMelin: #JustinTrudeau and Best Friends, #Obama, “president” #Joe Biden.  For
eign Policy: COVID MASKS, US DEMOCRATS and DOMINION VO…
## 2        #Biden Inauguration SWAG ...\nyou ain't all that #Joe,\nyou ain't legit.\n\n#maga @r
ealDonaldTrump @DonaldJTrumpJr… https://t.co/arQsbDolmE
## 3       Biden Picks Budget Director Who Pushed Social Security Cuts by @dailyposter : #NeeraT
andem is  a TERRIBLE choice, t… https://t.co/NA3kTJCoeA
## 4 RT @shrstraker: JESSICA TARLOV \nTHESE ARE THE CHILDREN \nOF THE POLITICAL #FAMILIES \nTHE
#SWAMP \nWHO SOLD OUT \nAMERICA TO BECOME #RICH \n#HUN…
## 5       #JoeBiden fractured his leg while playing with his dog ... looks like we’re going to
put #Joe down And make the dog… https://t.co/UP6Wvz3D7P
## 6     Antony Blinken and the tragedy of America’s foreign policy establishment\n\n@JohnHulsma
n1 on Antony Blinken as Americ… https://t.co/1umOfpIMJ0
##   favorited favoriteCount replyToSN             created truncated replyToSID
## 1     FALSE             0      <NA> 2020-11-30 20:33:48     FALSE       <NA>
## 2     FALSE             0      <NA> 2020-11-30 18:41:28      TRUE       <NA>
## 3     FALSE             0      <NA> 2020-11-30 17:31:28      TRUE       <NA>
## 4     FALSE             0      <NA> 2020-11-30 16:55:22     FALSE       <NA>
## 5     FALSE             2      <NA> 2020-11-30 14:38:36      TRUE       <NA>
## 6     FALSE             1      <NA> 2020-11-30 14:26:02      TRUE       <NA>
##                    id replyToUID
## 1 1333509557281198080       <NA>
## 2 1333481287332421632       <NA>
## 3 1333463673570902016       <NA>
## 4 1333454588465328133       <NA>
## 5 1333420167909560320       <NA>
## 6 1333417005798596614       <NA>
##                                                                          statusSource
## 1  <a href="http://twitter.com/#!/download/ipad" rel="nofollow">Twitter for iPad</a>
## 2            <a href="https://mobile.twitter.com" rel="nofollow">Twitter Web App</a>
## 3            <a href="https://mobile.twitter.com" rel="nofollow">Twitter Web App</a>
## 4            <a href="https://mobile.twitter.com" rel="nofollow">Twitter Web App</a>
## 5 <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone</a>
## 6                             <a href="https://buffer.com" rel="nofollow">Buffer</a>
##        screenName retweetCount isRetweet retweeted longitude latitude
## 1      BorjeMelin            2      TRUE     FALSE        NA       NA
## 2 2haveandhavenot            0     FALSE     FALSE        NA       NA
## 3  Stevevillano11            0     FALSE     FALSE        NA       NA
## 4      shrstraker            1      TRUE     FALSE        NA       NA
## 5         kinmack            1     FALSE     FALSE        NA       NA
## 6          CityAM            0     FALSE     FALSE        NA       NA
```

```
# Tiro só o que interessa que são os textos
electionUSA_text <- electionUSA$text
head(electionUSA_text)
```

```
## [1] "RT @BorjeMelin: #JustinTrudeau and Best Friends, #Obama, “president” #Joe Biden.  Foreig
n Policy: COVID MASKS, US DEMOCRATS and DOMINION VO…"
## [2] "#Biden Inauguration SWAG ...\nyou ain't all that #Joe,\nyou ain't legit.\n\n#maga @realD
onaldTrump @DonaldJTrumpJr… https://t.co/arQsbDolmE"
## [3] "Biden Picks Budget Director Who Pushed Social Security Cuts by @dailyposter : #NeeraTand
em is  a TERRIBLE choice, t… https://t.co/NA3kTJCoeA"
## [4] "RT @shrstraker: JESSICA TARLOV \nTHESE ARE THE CHILDREN \nOF THE POLITICAL #FAMILIES \nT
HE #SWAMP \nWHO SOLD OUT \nAMERICA TO BECOME #RICH \n#HUN…"
## [5] "#JoeBiden fractured his leg while playing with his dog ... looks like we’re going to put
#Joe down And make the dog… https://t.co/UP6Wvz3D7P"
## [6] "Antony Blinken and the tragedy of America’s foreign policy establishment\n\n@JohnHulsman
1 on Antony Blinken as Americ… https://t.co/1umOfpIMJ0"
```

```
# Faço o tratamento retirandos os caracteres e colocando tudo em minusculo.

# colocando em minusculo
electionUSA_text <- tolower(electionUSA_text)

# retirando os espações e brancos ('rt')
electionUSA_text <- gsub('rt', '', electionUSA_text)

# retirando os @ do indetificado do usuario
electionUSA_text <- gsub('@\\w+','',electionUSA_text)

# retirando pontuações
electionUSA_text <- gsub('[[:punct:]]','',electionUSA_text)

# retirando links nos textos
electionUSA_text <- gsub('http\\w+','',electionUSA_text)

# retirando os tabs dos textos
electionUSA_text <- gsub('[ |\t]{2,}','',electionUSA_text)

# retirando espaços em brancos no inicio
electionUSA_text <- gsub(' ^','',electionUSA_text)

# retirando espaços me branco no final
electionUSA_text <- gsub(' $','',electionUSA_text)

# retirando os emotions dos textos
electionUSA_text <- gsub("[^\x01-\x7F]", "", electionUSA_text)

head(electionUSA_text)
```

```
## [1] "justintrudeau and best friends obama president joe bidenforeign policy covid masks us de
mocrats and dominion vo"
## [2] "biden inauguration swag \nyou aint all that joe\nyou aint legit\n\nmaga"
## [3] "biden picks budget director who pushed social security cuts byneeratandem isa terrible c
hoice t"
## [4] "jessica tarlov \nthese are the children \nof the political families \nthe swamp \nwho so
ld out \namerica to become rich \nhun"
## [5] "joebiden fractured his leg while playing with his doglooks like were going to put joe do
wn and make the dog"
## [6] "antony blinken and the tragedy of americas foreign policy establishment\n\n on antony bl
inken as americ"
```

```
# Como o R trabalha com processamento em linguagem natural converto o vetor
# para o formato corpus.

corpus <- Corpus(VectorSource(electionUSA_text))
class(corpus)
```

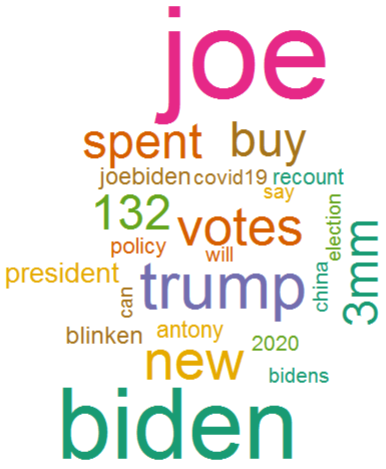
```
## [1] "SimpleCorpus" "Corpus"
```

```
electionUSA_corpus <- tm_map(corpus, function(x) removeWords(x, stopwords()))
```

```
## Warning in tm_map.SimpleCorpus(corpus, function(x) removeWords(x, stopwords())):
## transformation drops documents
```

```
# Grafico para mostrar as palavras mais usadas nos twitters

wordcloud(electionUSA_corpus, min.freq = 10, colors = brewer.pal(8,'Dark2'),
          random.color = TRUE, max.words = 1000)
```




```
# analyse de sentimentos, do pacote syuzhet
sentiment_election <- get_nrc_sentiment((electionUSA_text))

# Calcula o score para cada sentimento analisado e joga para um data frame.
sentiment_election_score <- data.frame(colSums(sentiment_election[,]))

# Atribuo um nome score para a soma dos scores de sentimentos.
names(sentiment_election_score) <- 'Score'
```

```
# Crio uma outra coluna com o nome sentiment pegando os nomes dos sentimentos pela função row.na
mes
# ficando assim com uma coluna para o sentimento e outra para a sua quantidade.
sentiment_election_score <- cbind('sentiment' = row.names(sentiment_election_score),sentiment_el
ection_score)

# Como criei a coluna sentimento retiro os nomes das linhas com os sentimentos
rownames(sentiment_election_score) <- NULL
head(sentiment_election_score)
```

```
##      sentiment Score
## 1        anger    40
## 2 anticipation    69
## 3      disgust    30
## 4         fear    48
## 5          joy    69
## 6      sadness    40
```

```
# Grafico de barra mostrando os sentimentos.
sentiment_election_score %>%
ggplot(aes(x = sentiment, y = Score, fill = sentiment)) +
  geom_bar(stat = "identity")+
  labs(x = 'Sentimento', y = 'Quantidade',
       title = 'Analise de sentimento eleição USA - Vitoria do Joe Biden',
       fill  = 'Sentimentos')
```