

Text Analytics Sentiment Analysis of Clients

Karim Souidi

September 16, 2019

Install Required Packages

```
options(repos=structure(c(CRAN="http://cran.utstat.utoronto.ca/"))))

install.packages(c("tidyverse", c("tidytext", dependencies =
TRUE), "devtools", "stringr", "magrittr",
                  "qdap", "tm", "readr", "stopwords", "SnowballC",
                  "wordcloud"))

## Warning: package 'TRUE' is not available (for R version 3.5.2)

## package 'tidyverse' successfully unpacked and MD5 sums checked
## package 'tidytext' successfully unpacked and MD5 sums checked
## package 'devtools' successfully unpacked and MD5 sums checked
## package 'stringr' successfully unpacked and MD5 sums checked
## package 'magrittr' successfully unpacked and MD5 sums checked
## package 'qdap' successfully unpacked and MD5 sums checked
## package 'tm' successfully unpacked and MD5 sums checked
## package 'readr' successfully unpacked and MD5 sums checked
## package 'stopwords' successfully unpacked and MD5 sums checked
## package 'SnowballC' successfully unpacked and MD5 sums checked
## package 'wordcloud' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\soudik\AppData\Local\Temp\RtmpIJ4XQk\downloaded_packages

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.5.3

## -- Attaching packages -----
## ----- tidyverse 1.2.1 -----

## v ggplot2 3.2.1      v purrr 0.3.2
## v tibble 2.1.3       v dplyr 0.8.3
## v tidyr 1.0.0        v stringr 1.4.0
## v readr 1.3.1       v forcats 0.4.0

## Warning: package 'ggplot2' was built under R version 3.5.3
## Warning: package 'tibble' was built under R version 3.5.3
## Warning: package 'tidyr' was built under R version 3.5.3
```

```
## Warning: package 'readr' was built under R version 3.5.3
## Warning: package 'purrr' was built under R version 3.5.3
## Warning: package 'dplyr' was built under R version 3.5.3
## Warning: package 'stringr' was built under R version 3.5.3
## Warning: package 'forcats' was built under R version 3.5.3
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(tidytext)

## Warning: package 'tidytext' was built under R version 3.5.3

library(devtools)

## Warning: package 'devtools' was built under R version 3.5.3
## Loading required package: usethis
## Warning: package 'usethis' was built under R version 3.5.3

library(stringr)
library(magrittr)

## Warning: package 'magrittr' was built under R version 3.5.3

##
## Attaching package: 'magrittr'

## The following object is masked from 'package:purrr':
##
##     set_names

## The following object is masked from 'package:tidyr':
##
##     extract

library(qdap)

## Warning: package 'qdap' was built under R version 3.5.3
## Loading required package: qdapDictionaries
## Loading required package: qdapRegex
## Warning: package 'qdapRegex' was built under R version 3.5.3

##
## Attaching package: 'qdapRegex'
```

```
## The following object is masked from 'package:dplyr':  
##  
##   explain  
  
## The following object is masked from 'package:ggplot2':  
##  
##   %>%  
  
## Loading required package: qdapTools  
  
## Warning: package 'qdapTools' was built under R version 3.5.3  
  
##  
## Attaching package: 'qdapTools'  
  
## The following object is masked from 'package:dplyr':  
##  
##   id  
  
## Loading required package: RColorBrewer  
  
##  
## Attaching package: 'qdap'  
  
## The following object is masked from 'package:magrittr':  
##  
##   %>%  
  
## The following object is masked from 'package:forcats':  
##  
##   %>%  
  
## The following object is masked from 'package:stringr':  
##  
##   %>%  
  
## The following object is masked from 'package:dplyr':  
##  
##   %>%  
  
## The following object is masked from 'package:purrr':  
##  
##   %>%  
  
## The following object is masked from 'package:tidyr':  
##  
##   %>%  
  
## The following object is masked from 'package:base':  
##  
##   Filter  
  
library\(tm\)
```

```
## Warning: package 'tm' was built under R version 3.5.3
## Loading required package: NLP
##
## Attaching package: 'NLP'
## The following object is masked from 'package:qdap':
##
##      ngrams
## The following object is masked from 'package:ggplot2':
##
##      annotate
##
## Attaching package: 'tm'
## The following objects are masked from 'package:qdap':
##
##      as.DocumentTermMatrix, as.TermDocumentMatrix

library(readr)
library(stopwords)

## Warning: package 'stopwords' was built under R version 3.5.3
##
## Attaching package: 'stopwords'
## The following object is masked from 'package:tm':
##
##      stopwords

library(SnowballC)
library(wordcloud)

## Warning: package 'wordcloud' was built under R version 3.5.3
```

Read Comments

```
text<-paste(readLines("K:/07 Surveys/03 Client/05 Email Survey/05
Responses/comments.csv"), collapse = "")
```

remove any dollar signs (they're special characters in R)

```
fileText <- gsub("\\$", "", text)
tokens <- data_frame(text = fileText) %>% unnest_tokens(word, text)

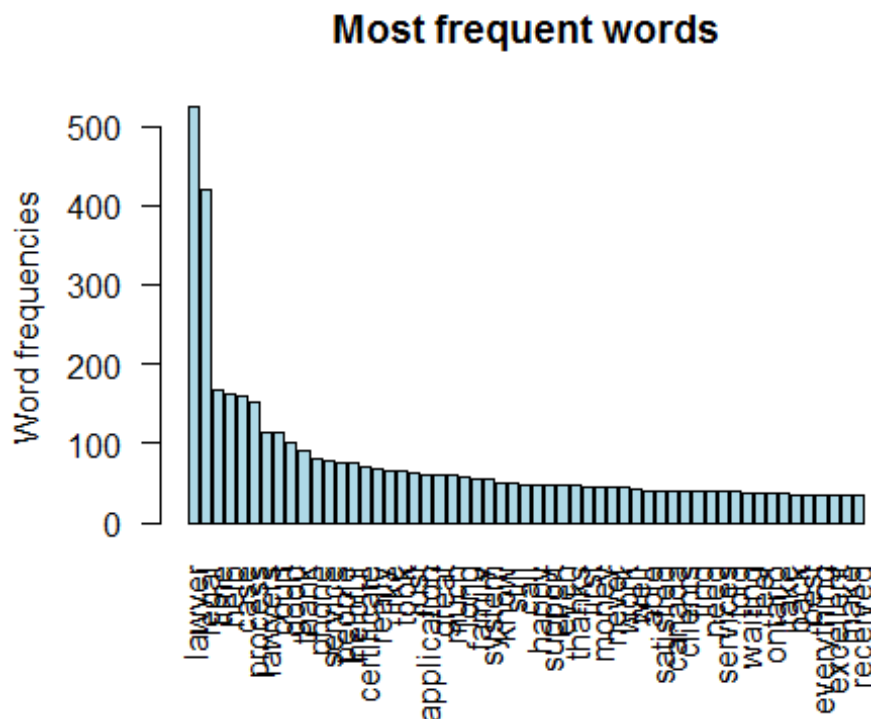
## Warning: `data_frame()` is deprecated, use `tibble()`.
## This warning is displayed once per session.

tokens %>%
  inner_join(get_sentiments("bing")) %>% # pull out only sentiment words
  count(sentiment) %>% # count the # of positive & negative words
```

```
## # A tibble: 1 x 3
##   negative positive sentiment
##   <dbl>     <dbl>     <dbl>
## 1     721     1587       866
```

```
freq_words<-freq_terms(tokens, top=50, at.least=4, stopwords("english"))
```

```
barplot(freq_words$FREQ, las = 2, names.arg = freq_words$WORD,
        col = "lightblue", main = "Most frequent words",
        ylab = "Word frequencies")
```



```
## Warning in tm_map.SimpleCorpus(corpus, tm::removePunctuation):  
## transformation drops documents
```

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

