

# Sentiment Analysis

Matthew Werner DAT 511 Bag of Words & Sentiment Analysis

```
require(ggplot2)

## Loading required package: ggplot2

x <- readLines("C:\\\\Users\\\\matth\\\\OneDrive\\\\Documents\\\\Grad School\\\\Data Analytics\\\\Spring 2022\\\\DAT511\\\\Reviews.txt")

head(x)

## [1] "Review"
## [2] " Everything from the weather, staff, food, property, fire pits, décor, spa, rooms and beach were great."
## [3] "\"The hotel it is fantastic built by the sea, living together with nature. Environment it is great and the service is excellent!\""
## [4] "We full enjoyed the place, and facilities."
## [5] "Thanks for the \"cidreira\" and \"madalenas\" tea at reception\""
## [6] "One dream! Cozy and comfortable Hotel! The best personalized one! All the employees are congratulatory and helpful"

library('corpus')

## Warning: package 'corpus' was built under R version 4.1.3

require(tm)

## Loading required package: tm

## Warning: package 'tm' was built under R version 4.1.3

## Loading required package: NLP

## 
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':
## 
##     annotate

sw_source<-VectorSource(x)

sw_corpus<-VCorpus(sw_source)
```

```

x<-tolower(x)
x<-removePunctuation(x)
x<-removeNumbers(x)
x<-removeWords(x, stopwords("en"))
x<-stripWhitespace(x)
clean_source<-VectorSource(x)
clean_corpus<-VCorpus(clean_source)

```

```

dtm<-TermDocumentMatrix(clean_corpus)
mterms<-as.matrix(dtm)
v<-sort(rowSums(mterms), decreasing=TRUE)
d<-data.frame(word=names(v), freq=v)
head(d, 20)

```

	word	freq
##		
## place	place	162
## thank	thank	148
## hotel	hotel	147
## wonderful	wonderful	108
## stay	stay	95
## room	room	90
## back	back	89
## staff	staff	86
## amazing	amazing	81
## experience	experience	79
## will	will	79
## one	one	78
## come	come	71
## food	food	71
## everything	everything	69
## beautiful	beautiful	65
## time	time	59
## nice	nice	57
## fantastic	fantastic	56
## great	great	56

```
require('wordcloud')
```

```

## Loading required package: wordcloud

## Warning: package 'wordcloud' was built under R version 4.1.3

```

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

```
wordcloud(words=d$word, freq=d$freq, min.freq=3, max.words=100, rot.per=0.35)
```

```

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 3, max.words =
## 100, : hotel could not be fit on page. It will not be plotted.

```

```

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 3, max.words =
## 100, : stay could not be fit on page. It will not be plotted.

```

```
## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 3, max.words =
## 100, : everything could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 3, max.words =
## 100, : reception could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 3, max.words =
## 100, : recommend could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 3, max.words =
## 100, : enjoyed could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = d$word, freq = d$freq, min.freq = 3, max.words =
## 100, : nice could not be fit on page. It will not be plotted.
```



Wordcloud insinuates that the reviews were positive.

```
findAssocs(dtm, terms="hotel", corlimit=0.2)
```

```

## $hotel
##      balanced knowledgeable      amazed      charmed      discover
##      0.28          0.28      0.27      0.27      0.27
##      erected      exciting      fits      funny      guest
##      0.27          0.27      0.27      0.27      0.27

```

```

##   incredibly      modest      mouth      offering      remains
##   0.27           0.27       0.27       0.27        0.27
##   restrained    specifically      surpasses      varied      played
##   0.27           0.27       0.27       0.27        0.26
##   relax          totally      products      rather      reception
##   0.26           0.26       0.24       0.24        0.24
##   restaurant     spa         friend      impressive      managed
##   0.24           0.24       0.23       0.23        0.23
##   size           compared      corners      decoration      either
##   0.23           0.22       0.22       0.22        0.22
##   enter          level       outdoor      sense      night
##   0.22           0.22       0.22       0.22        0.21
##   staff          without      bell        brief      changed
##   0.21           0.21       0.20       0.20        0.20
##   cleaners        confused      culminates      cut      desert
##   0.20           0.20       0.20       0.20        0.20
##   essentials      euros       finely      horta      hundreds
##   0.20           0.20       0.20       0.20        0.20
##   icing           loudly      meadow      midday      oclock
##   0.20           0.20       0.20       0.20        0.20
##   outdoors        outline      playing      practically      produce
##   0.20           0.20       0.20       0.20        0.20
##   rocking         series      six        travels      unnoticed
##   0.20           0.20       0.20       0.20        0.20
##   wastes
##   0.20

```

```
findAssocs(dtm,terms="staff",corlimit=0.2)
```

```

## $staff
##   spa receptionist      level      absolutely      charmed      cleaners
##   0.25           0.23       0.22       0.21        0.21       0.21
##   confused        crotchet      culminates      cut      desert      elevates
##   0.21           0.21       0.21       0.21        0.21       0.21
##   euros           finely      friendly      helpful      hotel      hundreds
##   0.21           0.21       0.21       0.21        0.21       0.21
##   keepers         loudly      meadow      midday      oclock      offering
##   0.21           0.21       0.21       0.21        0.21       0.21
##   outdoors        outline      playing      practically      promotes      series
##   0.21           0.21       0.21       0.21        0.21       0.21
##   six             smiley      strives      travels      unnoticed      varied
##   0.21           0.21       0.21       0.21        0.21       0.21
##   waitresses
##   0.21

```

```
findAssocs(dtm,terms="food",corlimit=0.2)
```

```

## $food
##   artist        bottle      experimenting      focusing      foreign
##   0.31           0.31       0.31       0.31        0.31       0.31
##   introduction      mini       pain        promote      show
##   0.31           0.31       0.31       0.31        0.31       0.31
##   clean          mattress      bad        taken      also

```

##	0.29	0.29	0.25	0.25	0.23
##	ground	leonardo	think	extremely	portuguese
##	0.23	0.23	0.23	0.22	0.22
##	three	arrive	beds	energy	evening
##	0.22	0.21	0.21	0.21	0.21
##	including	opportunity	addressing	adorable	afforded
##	0.21	0.21	0.20	0.20	0.20
##	appeared	appointed	awaits	blood	bonus
##	0.20	0.20	0.20	0.20	0.20
##	breast	celery	chicest	comes	command
##	0.20	0.20	0.20	0.20	0.20
##	comment	common	commune	components	consequences
##	0.20	0.20	0.20	0.20	0.20
##	constructive	contact	coolest	cultivated	dina
##	0.20	0.20	0.20	0.20	0.20
##	diners	distraction	dragged	effects	ensure
##	0.20	0.20	0.20	0.20	0.20
##	evenings	exotic	focused	furnishings	happier
##	0.20	0.20	0.20	0.20	0.20
##	harm	homes	ignorance	incident	johanna
##	0.20	0.20	0.20	0.20	0.20
##	knockon	locale	matters	merit	mistakes
##	0.20	0.20	0.20	0.20	0.20
##	networks	noma	nondescript	numerous	overwhelmed
##	0.20	0.20	0.20	0.20	0.20
##	partridge	partridges	passing	pigeon	placed
##	0.20	0.20	0.20	0.20	0.20
##	poisoning	puree	readjusting	recall	redoubled
##	0.20	0.20	0.20	0.20	0.20
##	refinement	remove	restaurant...	salmonella	security
##	0.20	0.20	0.20	0.20	0.20
##	seemingly	serious	similar	social	stained
##	0.20	0.20	0.20	0.20	0.20
##	stargaze	stood	strategically	strong	struck
##	0.20	0.20	0.20	0.20	0.20
##	stupidity	terms	throughout	tuesday	ultimately
##	0.20	0.20	0.20	0.20	0.20
##	unable	witness	youll		
##	0.20	0.20	0.20		

Hotel, Staff, and Food all are associated with words that describe their respective department.

```
require('SentimentAnalysis')
```

```
## Loading required package: SentimentAnalysis

## Warning: package 'SentimentAnalysis' was built under R version 4.1.3

##
## Attaching package: 'SentimentAnalysis'

## The following object is masked from 'package:base':
## 
##     write
```

```

require(SnowballC)

## Loading required package: SnowballC

require('SentimentAnalysis')
sa<-analyzeSentiment(x)
head(sa)

##   WordCount SentimentGI NegativityGI PositivityGI SentimentHE NegativityHE
## 1          1 0.00000000 0.00000000 0.0000000 0.00000000          0
## 2          13 -0.07692308 0.07692308 0.0000000 0.00000000          0
## 3          12 0.41666667 0.08333333 0.5000000 0.00000000          0
## 4          4 0.50000000 0.00000000 0.5000000 0.25000000          0
## 5          5 0.40000000 0.00000000 0.4000000 0.00000000          0
## 6         35 0.20000000 0.08571429 0.2857143 0.02857143          0
##   PositivityHE SentimentLM NegativityLM PositivityLM RatioUncertaintyLM
## 1 0.00000000 0.00000000 0.00000000 0.00000000          0
## 2 0.00000000 -0.07692308 0.07692308 0.00000000          0
## 3 0.00000000 0.16666667 0.00000000 0.16666667          0
## 4 0.25000000 0.25000000 0.00000000 0.25000000          0
## 5 0.00000000 0.20000000 0.00000000 0.20000000          0
## 6 0.02857143 0.02857143 0.05714286 0.08571429          0
##   SentimentQDAP NegativityQDAP PositivityQDAP
## 1 0.00000000 0.00000000 0.00000000
## 2 0.07692308 0.00000000 0.07692308
## 3 0.41666667 0.00000000 0.41666667
## 4 0.25000000 0.00000000 0.25000000
## 5 0.40000000 0.00000000 0.40000000
## 6 0.20000000 0.02857143 0.22857143

head(convertToBinaryResponse(sa)$SentimentQDAP)

## [1] positive positive positive positive positive positive
## Levels: negative positive

dict.QDAP<-loadDictionaryQDAP()
dict.GI<-loadDictionaryGI()

dtemp<-preprocessCorpus(clean_corpus)
sentiment_direction<-convertToDirection(analyzeSentiment(dtemp)$SentimentQDAP)

for (k in 1:10)
{
  print(sentiment_direction[k])
  cat(x[k], "\n")
}

## [1] neutral
## Levels: negative neutral positive
## review

```

```

## [1] positive
## Levels: negative neutral positive
##   everything weather staff food property fire pits décor spa rooms beach top notch
## [1] positive
## Levels: negative neutral positive
##   hotel fantastic built sea living together nature environment great well people service
## [1] positive
## Levels: negative neutral positive
##   full enjoyed place facilities
## [1] positive
## Levels: negative neutral positive
##   thanks cidreira madalenas tea reception
## [1] positive
## Levels: negative neutral positive
##   one dream cozy comfortable hotel best personalized one employees congratulated nice taste since rece
## [1] positive
## Levels: negative neutral positive
##   hotel concept hard grasp communicate environmental sustainability fashionable buzzwords like upcyclin
## [1] negative
## Levels: negative neutral positive
##   pricing utterly ridiculous get
## [1] neutral
## Levels: negative neutral positive
##   breakfast highquality products lacks choice
## [1] positive
## Levels: negative neutral positive
##   service friendly although friendly times

dtemp<-preprocessCorpus(sw_corpus)
sw_sa<-analyzeSentiment(dtemp)
sentiment_direction<-convertToDirection(sw_sa$SentimentQDAP)

str(sw_sa)

## 'data.frame': 652 obs. of 14 variables:
## $ WordCount      : num  1 13 12 4 5 35 16 4 5 5 ...
## $ SentimentGI    : num  0 -0.0769 0.4167 0.5 0.4 ...
## $ NegativityGI   : num  0 0.0769 0.0833 0 0 ...
## $ PositivityGI   : num  0 0 0.5 0.5 0.4 ...
## $ SentimentHE    : num  0 0 0 0.25 0 ...
## $ NegativityHE   : num  0 0 0 0 0 0 0 0 0 ...
## $ PositivityHE   : num  0 0 0 0.25 0 ...
## $ SentimentLM    : num  0 -0.0769 0.1667 0.25 0.2 ...
## $ NegativityLM   : num  0 0.0769 0 0 0 ...
## $ PositivityLM   : num  0 0 0.167 0.25 0.2 ...
## $ RatioUncertaintyLM: num  0 0 0 0 0 0 0 0 0 ...
## $ SentimentQDAP   : num  0 0.0769 0.4167 0.25 0.4 ...
## $ NegativityQDAP  : num  0 0 0 0 0 0 0.0625 0.5 0.2 0 ...
## $ PositivityQDAP  : num  0 0.0769 0.4167 0.25 0.4 ...

for (k in 1:10)
{
  print(sentiment_direction[k])

```

```

print(unlist(sw_corpus[k])["content.content"])
cat("\n")
}

## [1] neutral
## Levels: negative neutral positive
## content.content
## "Review"
##
## [1] positive
## Levels: negative neutral positive
## content
## " Everything from the weather, staff, food, property, fire pits, décor, spa, rooms and beach were top
##
## [1] positive
## Levels: negative neutral positive
##
## "The hotel it is fantastic built by the sea, living together with nature. Environment it is great a
##
## [1] positive
## Levels: negative neutral positive
## content.content
## "We full enjoyed the place, and facilities."
##
## [1] positive
## Levels: negative neutral positive
## content.content
## "Thanks for the \"cidreira\" and \"madalenas\" tea at reception\""
##
## [1] positive
## Levels: negative neutral positive
##
## "One dream! Cozy and comfortable Hotel! The best personalized one! All the employees are congratula
##
## [1] positive
## Levels: negative neutral positive
##
## "Hotel concept is hard to grasp. They communicate environmental sustainability, through fashionabl
##
## [1] negative
## Levels: negative neutral positive
## content.content
## "Pricing is utterly ridiculous, for what you get."
##
## [1] neutral
## Levels: negative neutral positive
## content.content
## "Breakfast has high-quality products, but it lacks in choice."
##
## [1] positive
## Levels: negative neutral positive
## content.content
## "Service is friendly, although \"too friendly\" at times."

```

```

table(convertToDirection(sw_sa$SentimentQDAP))

##
## negative  neutral  positive
##      17        57       560

table(convertToDirection(sw_sa$SentimentLM))

##
## negative  neutral  positive
##      88        245       301

unlist(sw_corpus[which.max(sw_sa$SentimentQDAP)])["content.content"]

##                               content.content
## "Thanks, 1000 thanks for this again... "

sort((sw_sa$SentimentQDAP),decreasing=F)[1:10]

## [1] -0.50000000 -0.25000000 -0.21428571 -0.20000000 -0.16666667 -0.14285714
## [7] -0.13157895 -0.12000000 -0.11111111 -0.09090909

unlist(sw_corpus[which.min(sw_sa$SentimentQDAP)])["content.content"]

##                               content.content
## "Pricing is utterly ridiculous, for what you get."

cor(na.omit(sw_sa[, c("SentimentLM", "SentimentHE", "SentimentQDAP", "SentimentGI")]))


##          SentimentLM SentimentHE SentimentQDAP SentimentGI
## SentimentLM    1.0000000   0.4067972   0.3401876   0.2667891
## SentimentHE    0.4067972   1.0000000   0.1372350   0.1122909
## SentimentQDAP   0.3401876   0.1372350   1.0000000   0.8091213
## SentimentGI    0.2667891   0.1122909   0.8091213   1.0000000

```

The correlation between GI and QDAP is relatively high, which would be expected.