Hotel Review Analysis

LIL

September 21, 2018

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

#search.string <- readline(prompt="Enter an Text: ")  
#install.packages("tm",dependencies=TRUE)  
#install.packages("NLP",dependencies=TRUE)  
library("tm")

## Warning: package 'tm' was built under R version 3.4.4

## Loading required package: NLP

rw=read.csv('HotelRw\_csv.csv')  
  
rw\_d=data.frame(rw)  
library(tm)  
myCorpus <- Corpus(VectorSource(rw\_d$Review))  
#myCorpus <- tm\_map(myCorpus, content\_transformer(tolower))  
removeURL <- function(x) gsub("http[^[:space:]]\*", "", x)  
  
# remove URLs  
  
removeURL <- function(x) gsub("http[^[:space:]]\*", "", x)  
  
#myCorpus<-Corpus(VectorSource(tolower(myCorpus)))  
  
myCorpus <- tm\_map(myCorpus, content\_transformer(removeURL))

## Warning in tm\_map.SimpleCorpus(myCorpus, content\_transformer(removeURL)):  
## transformation drops documents

# remove anything other than English letters or space  
  
removeNumPunct <- function(x) gsub("[^[:alpha:][:space:]]\*", "", x)  
  
myCorpus <- tm\_map(myCorpus, content\_transformer(removeNumPunct))

## Warning in tm\_map.SimpleCorpus(myCorpus,  
## content\_transformer(removeNumPunct)): transformation drops documents

myCorpus <- tm\_map(myCorpus, content\_transformer(tolower))

## Warning in tm\_map.SimpleCorpus(myCorpus, content\_transformer(tolower)):  
## transformation drops documents

# remove stopwords  
  
  
#myCorpus <- tm\_map(myCorpus, content\_transformer(tolower))  
#stopwordlist<-c("r", "big","use", "see", "used", "via","https", "amp","will","this","also","first","call","till","cant","like","what","This","THIS")  
#line=gsub("#","",search.string)  
#write(line,file="D:/sales force/tweeter analysis/stopWords.csv",append=TRUE)  
  
#stopwordlist<-read.csv(choose.files(),header = FALSE,sep = ",")  
  
stopwordlist<-read.csv(file="D:/sales force/tweeter analysis/stopWords.csv",header = FALSE,sep = ",")  
  
myStopwords <- c(setdiff(stopwords('english'), c("r", "big")),"use", "see", "used", "via","https", "amp","will","this","also","first","modi","rahul","jhasanjay")  
  
#########  
########  
#myStopwords <- c(setdiff(stopwords('english'),search\_text))  
  
myStopwords1<-unlist(myStopwords)  
  
stopwordlist<-unlist(stopwordlist)  
  
myCorpus <- tm\_map(myCorpus, removeWords, myStopwords1)

## Warning in tm\_map.SimpleCorpus(myCorpus, removeWords, myStopwords1):  
## transformation drops documents

#myCorpus <- tm\_map(myCorpus, removeWords, stopwordlist)  
  
  
#myStopwords <- c(stopwords('english'), stopwordlist)  
#myStopwords<-unlist(myStopwords)  
  
  
#myStopwords<-unlist(myStopwords)  
myCorpus <- tm\_map(myCorpus, removeWords, myStopwords)

## Warning in tm\_map.SimpleCorpus(myCorpus, removeWords, myStopwords):  
## transformation drops documents

#myCorpus <- tm\_map(myCorpus, removeWords, "khairatabad")  
  
myCorpus <- tm\_map(myCorpus, removeWords, c("this","will"))

## Warning in tm\_map.SimpleCorpus(myCorpus, removeWords, c("this", "will")):  
## transformation drops documents

myCorpus <- tm\_map(myCorpus, removeWords, c("must","Must","MUST","bjp","Bjp","BJP"))

## Warning in tm\_map.SimpleCorpus(myCorpus, removeWords, c("must", "Must", :  
## transformation drops documents

# remove extra whitespace  
  
  
myCorpus <- tm\_map(myCorpus, stripWhitespace)

## Warning in tm\_map.SimpleCorpus(myCorpus, stripWhitespace): transformation  
## drops documents

# keep a copy for stem completion later  
  
myCorpusCopy <- myCorpus  
  
##To Avoid incomplete words dnt run following command  
  
#docs <- tm\_map(docs, PlainTextDocument) # needs to come before stemming  
#docs <- tm\_map(docs, stemDocument, "english")  
  
  
  
#myCorpus <- tm\_map(myCorpus, stemDocument)  
  
head(myCorpus)

## <<SimpleCorpus>>  
## Metadata: corpus specific: 1, document level (indexed): 0  
## Content: documents: 6

#install.packages("SnowballC")  
library(SnowballC)  
##Minimum word length is 3 charactes  
tdm <- TermDocumentMatrix(myCorpus,control = list(wordLengths = c(4, Inf)))  
  
#head(tdm)  
#install.packages("wordcloud")  
  
library(wordcloud)

## Warning: package 'wordcloud' was built under R version 3.4.4

## Loading required package: RColorBrewer

m <- as.matrix(tdm)  
  
# calculate the frequency of words and sort it by frequency  
  
  
word.freq <- sort(rowSums(m), decreasing = T)  
  
#wordcloud(words = names(word.freq), freq = word.freq,scale=c(4,.5), min.freq = 5, random.order = F)  
  
wordcloud(words = names(word.freq), freq = word.freq,scale=c(4,.5), min.freq = 5, random.order = F,colors=brewer.pal(6, "Dark2"))

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : absolutely could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : impressed could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : tasted could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : cooked could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : wonderful could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : nothing could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : eating could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : ambiance could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : seriously could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : area could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : chips could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : friends could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : family could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : super could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : hour could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : poor could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : shrimp could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : tender could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : horrible could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : check could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : mediocre could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : management could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : done could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : small could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : either could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : special could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : money could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : authentic could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : dinner could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : kept could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : asked could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : tables could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : quick could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : portions could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : dessert could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : outside could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : pasta could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : customer could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : sweet could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : disappointing could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : thai could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : eaten could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : places could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : away could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : thought could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : extremely could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : seated could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : avoid could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : bacon could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : zero could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : bread could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : things could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : waiting could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : owners could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : wrong could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : fried could not be fit on page. It will not be plotted.

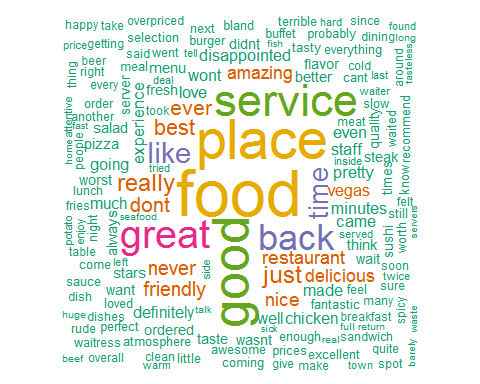
## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : expect could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : reasonable could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : enjoyed could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : anytime could not be fit on page. It will not be plotted.

## Warning in wordcloud(words = names(word.freq), freq = word.freq, scale =  
## c(4, : bring could not be fit on page. It will not be plotted.



## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.