# WSMA Project

Suprasanna Pradhan

31 August 2019

# 1. Project Objective and Scope

## 1.1 Objective

Shark Tank - the reality show where entrepreneurs pitch ideas to a panel of celebrity A dataset of Shark Tank episodes is made available. It contains 495 entrepreneurs making their pitch to the VC sharks. We will only use “Description” column for the initial text mining exercise.

Looking back at all of this got us curious: clasifying Shark deal? Are certain deals bheaviour more likely to make deals than others? What types of products are the most successful historically and recored in the description ? Which sharks dealthrow down the largest ration.

## 1.2 Scope

Using multiple algorithms, we will predict given the description of new pitch, how likely is the pitch will convert into success or not.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tm)
```

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

```
library(wordcloud)
```

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

```
library(plotrix)
library(dendextend)
```

```
##
## ---------------------
## Welcome to dendextend version 1.12.0
## Type citation('dendextend') for how to cite the package.
##
## Type browseVignettes(package = 'dendextend') for the package vignette.
## The github page is: https://github.com/talgalili/dendextend/
##
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dende
xtend/issues
## Or contact: <tal.galili@gmail.com>
##
##  To suppress this message use:  suppressPackageStartupMessages(library(dendextend))
## ---------------------
```

```
##
## Attaching package: 'dendextend'
```

```
## The following object is masked from 'package:stats':
##
##     cutree
```

```
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
```

```
## The following object is masked from 'package:NLP':
##
##     annotate
```

```
library(ggthemes)
library(reshape2)
library(quanteda)
```

```
## Package version: 1.5.1
```

```
## Parallel computing: 2 of 4 threads used.
```

```
## See https://quanteda.io for tutorials and examples.
```

```
##
## Attaching package: 'quanteda'
```

```
## The following objects are masked from 'package:tm':
##
##     as.DocumentTermMatrix, stopwords
```

```
## The following object is masked from 'package:utils':
##
##     View
```

```
library(stringr)
library(tidytext)
library(tidyr)
```

```
##
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:reshape2':
##
##     smiths
```

```
library(SnowballC)
library(irlba)
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following object is masked from 'package:tidyr':
##
##     expand
```

```
library(caret)
```

```
## Loading required package: lattice
```

```
library(RColorBrewer)
library(biclust)
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
## Loading required package: grid
```

```
## Loading required package: colorspace
```

```
library(igraph)
```

```
##
## Attaching package: 'igraph'
```

```
## The following object is masked from 'package:tidyr':
##
##     crossing
```

```
## The following object is masked from 'package:quanteda':
##
##     as.igraph
```

```
## The following objects are masked from 'package:dplyr':
##
##     as_data_frame, groups, union
```

```
## The following objects are masked from 'package:stats':
##
##     decompose, spectrum
```

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

```
library(fpc)
```

# 2. Project Approach

A typical Development Lifecycle can be adopted for this assignment, as follows: 1. Discovery and data cleaning 2. Data Preparation 3. Planning of Model 4. Building of Model 5. Final Results.

# 2.1. Discovery and data cleaning

## Import Dataset

```
#Importing Data set
setwd("C:/Users/SuprasannaPradhan/Documents/My Files/Great Lakes Projects")
Shark=read.csv("S_T_Companies.csv",stringsAsFactors = FALSE)
names(Shark)
```

```
##  [1] "deal"                  "description"
##  [3] "episode"               "category"
##  [5] "entrepreneurs"         "location"
##  [7] "website"               "askedFor"
##  [9] "exchangeForStake"      "valuation"
## [11] "season"                "shark1"
## [13] "shark2"                "shark3"
## [15] "shark4"                "shark5"
## [17] "title"                 "episode.season"
## [19] "Multiple.Entreprenuers"
```

```
str(Shark)
```

```
## 'data.frame':    495 obs. of  19 variables:
##  $ deal                  : logi  FALSE TRUE TRUE FALSE FALSE TRUE ...
##  $ description           : chr  "Bluetooth device implant for your ear." "Retail an
d wholesale pie factory with two retail locations in New Jersey." "Ava the Elephant is
a godsend for frazzled parents of young children everywhere. This talking medicine dis
pense"| __truncated__ "Organizing, packing, and moving services delivered by college w
omen." ...
##  $ episode               : int  1 1 1 1 1 2 2 2 2 2 ...
##  $ category              : chr  "Novelties" "Specialty Food" "Baby and Child Care"
"Consumer Services" ...
##  $ entrepreneurs         : chr  "Darrin Johnson" "Tod Wilson" "Tiffany Krumins" "Ni
ck Friedman, Omar Soliman" ...
##  $ location              : chr  "St. Paul, MN" "Somerset, NJ" "Atlanta, GA" "Tampa,
FL" ...
##  $ website               : chr  "" "http://whybake.com/" "http://www.avatheelephan
t.com/" "http://collegehunkshaulingjunk.com/" ...
##  $ askedFor              : int  1000000 460000 50000 250000 1200000 500000 200000 1
00000 500000 250000 ...
##  $ exchangeForStake      : int  15 10 15 25 10 15 20 20 10 10 ...
##  $ valuation             : int  6666667 4600000 333333 1000000 12000000 3333333 100
0000 500000 5000000 2500000 ...
##  $ season                : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ shark1                : chr  "Barbara Corcoran" "Barbara Corcoran" "Barbara Corc
oran" "Barbara Corcoran" ...
##  $ shark2                : chr  "Robert Herjavec" "Robert Herjavec" "Robert Herjave
c" "Robert Herjavec" ...
##  $ shark3                : chr  "Kevin O'Leary" "Kevin O'Leary" "Kevin O'Leary" "Ke
vin O'Leary" ...
##  $ shark4                : chr  "Daymond John" "Daymond John" "Daymond John" "Daymo
nd John" ...
##  $ shark5                : chr  "Kevin Harrington" "Kevin Harrington" "Kevin Harrin
gton" "Kevin Harrington" ...
##  $ title                 : chr  "Ionic Ear" "Mr. Tod's Pie Factory" "Ava the Elepha
nt" "College Foxes Packing Boxes" ...
##  $ episode.season        : chr  "01-Jan" "01-Jan" "01-Jan" "01-Jan" ...
##  $ Multiple.Entreprenuers: logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
```

```
table(Shark$deal)
```

```
##
## FALSE  TRUE
##   244   251
```

Above we got 495 observations with 19 variables, here we are going use only deal and desrcipaiton colmuon for text analyis Deal contains class of true and flase ,where 244 flase and 251 trues observations

# Create corpus

Further we need to transform dataset into a corpus with required variable i.e. description. Next we normalize the texts in the reviews: 1. Switch to lower case 2. Remove punctuation marks and stopwords 3. Remove extra whitespaces 4. Stem the documents

```
#Make a vector source and a corpus
Shark_corpus=Corpus(VectorSource(Shark$description))
```

# Clean the corpus and convert all text to lower case

# remove numbers$punctuatio$ stop words

```
#clean the corpus and convert all text to lower case
Shark_clean_corpus = tm_map(Shark_corpus, tolower)
```

```
## Warning in tm_map.SimpleCorpus(Shark_corpus, tolower): transformation drops
## documents
```

```
#remove numbers
Shark_clean_corpus = tm_map(Shark_clean_corpus,removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(Shark_clean_corpus, removeNumbers):
## transformation drops documents
```

```
# remove punctuation
Shark_clean_corpus = tm_map(Shark_clean_corpus,removePunctuation)
```

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

```
# remove whitespace
Shark_clean_corpus = tm_map(Shark_clean_corpus,stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(Shark_clean_corpus, stripWhitespace):
## transformation drops documents
```

```
# remove stop words
Shark_clean_corpus = tm_map(Shark_clean_corpus, removeWords, stopwords('en'))
```

```
## Warning in tm_map.SimpleCorpus(Shark_clean_corpus, removeWords,
## stopwords("en")): transformation drops documents
```

```
# Remove context specific stop words
Shark_clean_corpus=tm_map(Shark_clean_corpus, removeWords,c("also", "get","like", "com
pany", "made", "can", "im", "dress", "just", "i","3d"))
```

```
## Warning in tm_map.SimpleCorpus(Shark_clean_corpus, removeWords, c("also", :
## transformation drops documents
```

```
# stemming
Shark_clean_corpus = tm_map(Shark_clean_corpus, stemDocument)
```

```
## Warning in tm_map.SimpleCorpus(Shark_clean_corpus, stemDocument):
## transformation drops documents
```

```
Shark_clean_corpus
```

```
## <<SimpleCorpus>>
## Metadata:  corpus specific: 1, document level (indexed): 0
## Content:  documents: 495
```

## Creat word cloud

```
#Creat word cloud
wordcloud(Shark_clean_corpus, min.freq =10)
```

```
## Warning in wordcloud(Shark_clean_corpus, min.freq = 10): make could not be
## fit on page. It will not be plotted.
```


```
wordcloud(Shark_clean_corpus, min.freq =10, color = brewer.pal(8, "Set2"), random.orde
r=F, rot.per=.30)
```


```
wordcloud(Shark_clean_corpus, min.freq =10, color = brewer.pal(8, "Set2"), random.orde
r=F, rot.per=.30)
```


Above observation we found “make”" is the most frequest word has been appred several time

# Sentiment Analysis

```
#Sentimentr#
library(sentimentr)
library(syuzhet)
```

```
##
## Attaching package: 'syuzhet'
```

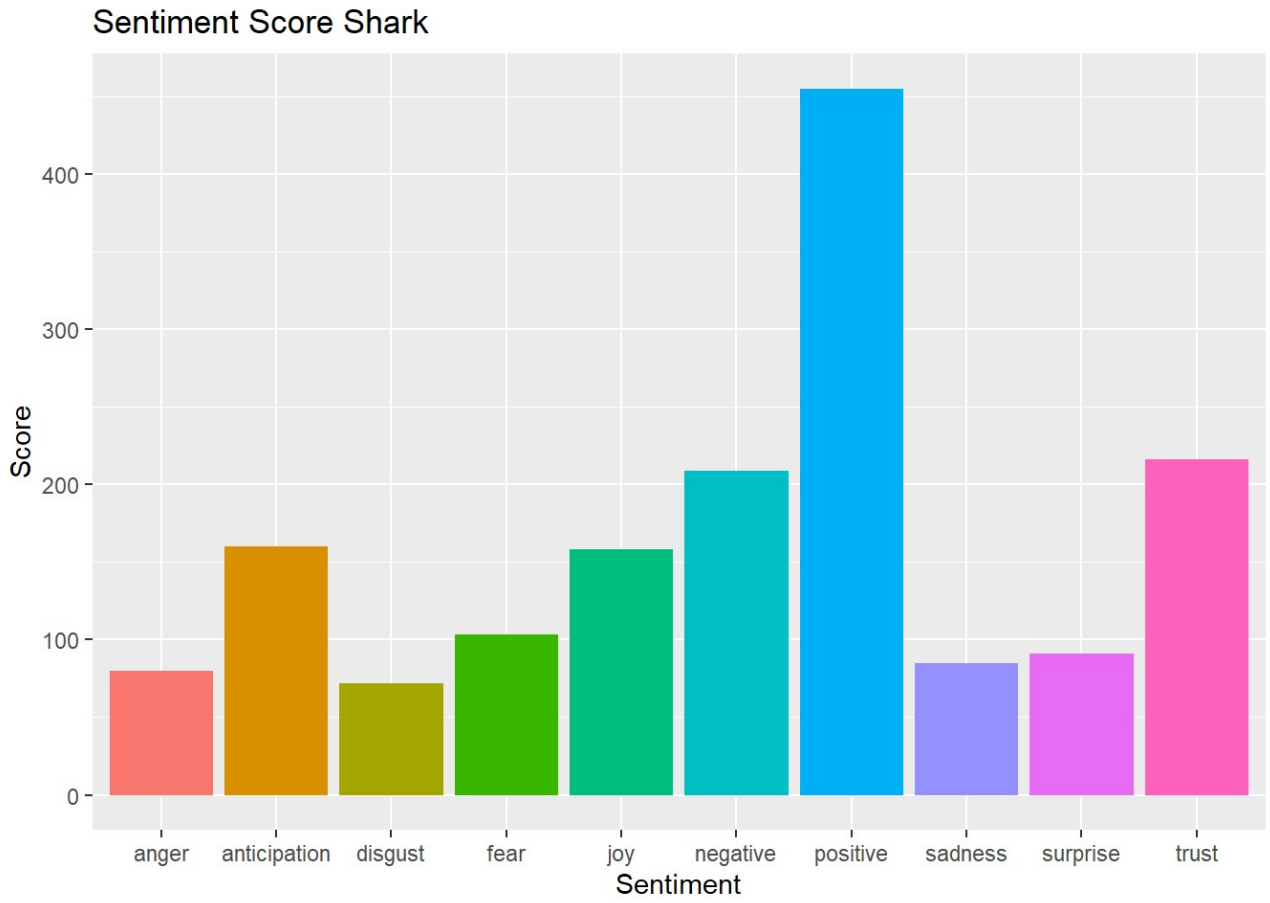
```
## The following object is masked from 'package:sentimentr':
##
##     get_sentences
```

```
## The following object is masked from 'package:plotrix':
##
##     rescale
```

```
library(ggplot2)
txt1 = gsub("(|via)((?:\\b\\W*@\\w+)+)", " ", Shark_corpus)
#Remove RT text etc
txt2 = gsub("http[^[:blank:]]+", " ", txt1)
#Remove html
txt3 = gsub("@\\w+", "", txt2)
#Remove names
txt4 = gsub("[[:punct:]]", " ", txt3)
txt5 = gsub("[^[:alnum:]]", " ", txt4)
```

We have removed all unnecessary words and letter whihc is notrequried for our anlysis

```
IXRsentiment = get_nrc_sentiment(txt5)
SentimentScore = data.frame(colSums(IXRsentiment[,]))
names(SentimentScore) = "Score"
SentimentScore = cbind("sentiment" = rownames(SentimentScore), SentimentScore)
rownames(SentimentScore) = NULL
ggplot(data = SentimentScore, aes(x = sentiment, y = Score))+
  geom_bar(aes(fill=sentiment), stat = "identity") +
  theme(legend.position = "none") +
  xlab("Sentiment") + ylab("Score") + ggtitle("Sentiment Score Shark")
```




###Genrate a document term matrix and finding association s DTM (Document-Term Matrix):the documents as rows, terms/words as columns, frequency of the term in the document. This will help us identify unique words in the corpus used frequently.

```
# Genrate a document term matrix
DTM =DocumentTermMatrix(Shark_clean_corpus)
DTM
```

```
## <<DocumentTermMatrix (documents: 495, terms: 3459)>>
## Non-/sparse entries: 9138/1703067
## Sparsity           : 99%
## Maximal term length: 21
## Weighting          : term frequency (tf)
```

```
#findFreqTerms(DTM, lowfreq=5)
findFreqTerms(DTM, lowfreq=10)
```

```
##   [1] "devic"     "new"       "retail"    "two"       "children"
##   [6] "easi"      "make"      "one"       "parent"    "play"
##  [11] "provid"    "turn"      "deliv"     "organ"     "servic"
##  [16] "women"     "offer"     "design"    "first"     "flavor"
##  [21] "food"      "includ"    "line"      "mani"      "perfect"
##  [26] "product"   "sold"      "store"     "activ"     "apparel"
##  [31] "brand"     "cloth"     "look"      "mix"       "attach"
##  [36] "hold"      "start"     "vehicl"    "learn"     "work"
##  [41] "item"      "find"      "help"      "fit"       "card"
##  [46] "combin"    "fun"       "keep"      "kid"       "remov"
##  [51] "add"       "babi"      "easier"    "protect"   "quick"
##  [56] "size"      "solut"     "time"      "yet"       "coffe"
##  [61] "toy"       "use"       "back"      "buy"       "gift"
##  [66] "sell"      "year"      "accessori" "color"     "onlin"
##  [71] "user"      "bar"       "enjoy"     "ingredi"   "market"
##  [76] "natur"     "safe"      "sugar"     "well"      "famili"
##  [81] "fashion"   "base"      "bodi"      "custom"    "high"
##  [86] "person"    "bottl"     "around"    "materi"    "featur"
##  [91] "live"      "instead"   "three"     "belt"      "uniqu"
##  [96] "allow"     "special"   "dog"       "need"      "cover"
## [101] "full"      "run"       "allnatur"  "blend"     "patent"
## [106] "place"     "clean"     "easili"    "pocket"    "way"
## [111] "busi"      "cream"     "ice"       "mobil"     "even"
## [116] "take"      "your"      "creat"     "shape"     "usa"
## [121] "give"      "want"      "shoe"      "bag"       "men"
## [126] "money"     "plastic"   "power"     "produc"    "tie"
## [131] "travel"    "without"   "cup"       "premium"   "wine"
## [136] "fresh"     "air"       "bring"     "come"      "effect"
## [141] "reduc"     "ride"      "tradit"    "serv"      "avail"
## [146] "friend"    "anyon"     "pet"       "smell"     "spray"
## [151] "truck"     "small"     "skin"      "builtin"   "cake"
## [156] "pad"       "system"    "train"     "game"      "altern"
## [161] "compani"   "fire"      "home"      "tool"      "box"
## [166] "kit"       "toilet"    "contain"   "water"     "peopl"
## [171] "bed"       "transform" "day"       "music"     "open"
## [176] "real"      "light"     "butter"    "free"      "oil"
## [181] "packag"    "name"      "now"       "chang"     "better"
## [186] "comfort"   "best"      "hand"      "varieti"   "treat"
## [191] "balm"      "everi"     "rang"      "wast"      "drink"
## [196] "put"       "wheel"     "chocol"    "nut"       "smartphon"
## [201] "app"       "style"     "phone"     "afford"    "dont"
## [206] "hair"      "sound"     "simpl"     "great"
```

```
findAssocs(DTM, "make",0.25)
```

```
## $make
##  easier cracker  strike   spong    imag
##    0.31    0.29    0.29    0.25    0.25
```

```
m = as.matrix(DTM)
v = sort(rowSums(m),decreasing=TRUE)
d = data.frame(word = names(v),freq=v)
head(d,10)
```

```
##     word freq
## 443  443  109
## 59    59   84
## 379  379   83
## 179  179   78
## 434  434   77
## 368  368   76
## 65    65   74
## 433  433   74
## 129  129   73
## 414  414   73
```

In above since words text are not appred we have to do something alternative Further we have prepared also the TDM (TDM is DTM (document term matrix) where words in TDM swap positions to constitute a DTM.)

```
#Genrate a term document matrix(Terms - Row, document - Columns)#
TDM = TermDocumentMatrix(Shark_clean_corpus)
class(TDM)
```

```
## [1] "TermDocumentMatrix"    "simple_triplet_matrix"
```

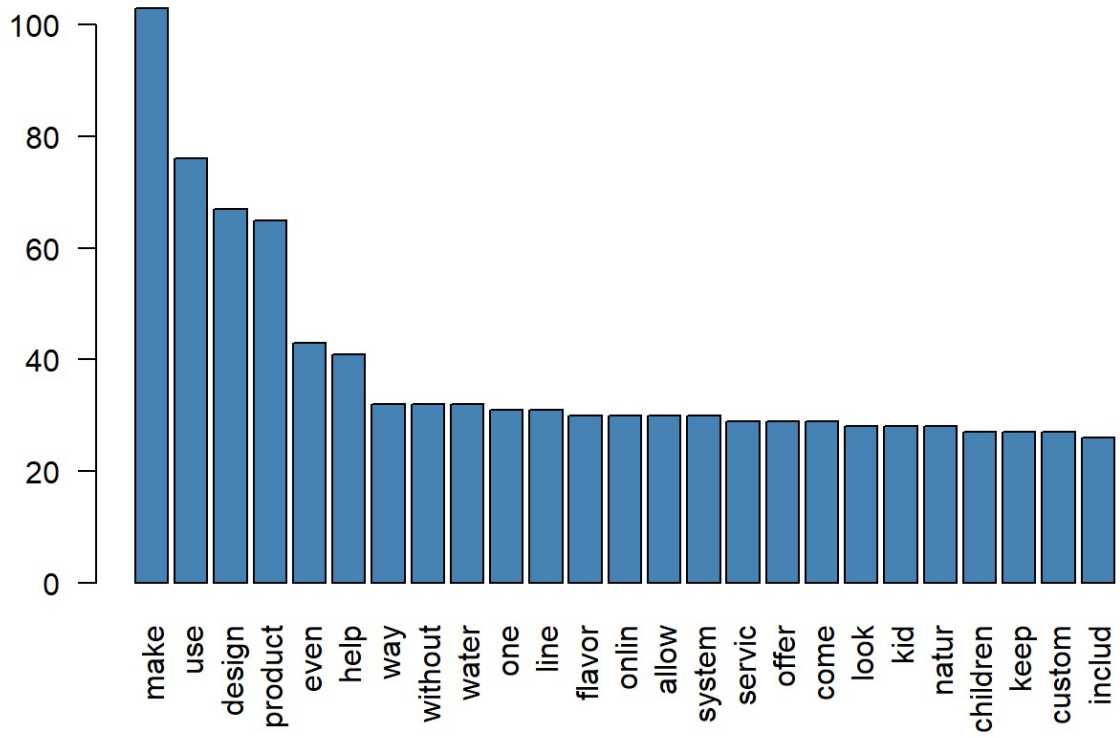
## Convert TDM to a matrix object

```
#Convert tdm to a matrix object #
m1 = as.matrix(TDM)
v1 = sort(rowSums(m1),decreasing=TRUE)
d1 = data.frame(word = names(v1),freq=v1)
head(d1,10)
```

```
##            word freq
## make       make  103
## use         use   76
## design   design   67
## product product   65
## even       even   43
## help       help   41
## way         way   32
## without without   32
## water     water   32
## one         one   31
```

Above words are having maximum frequences

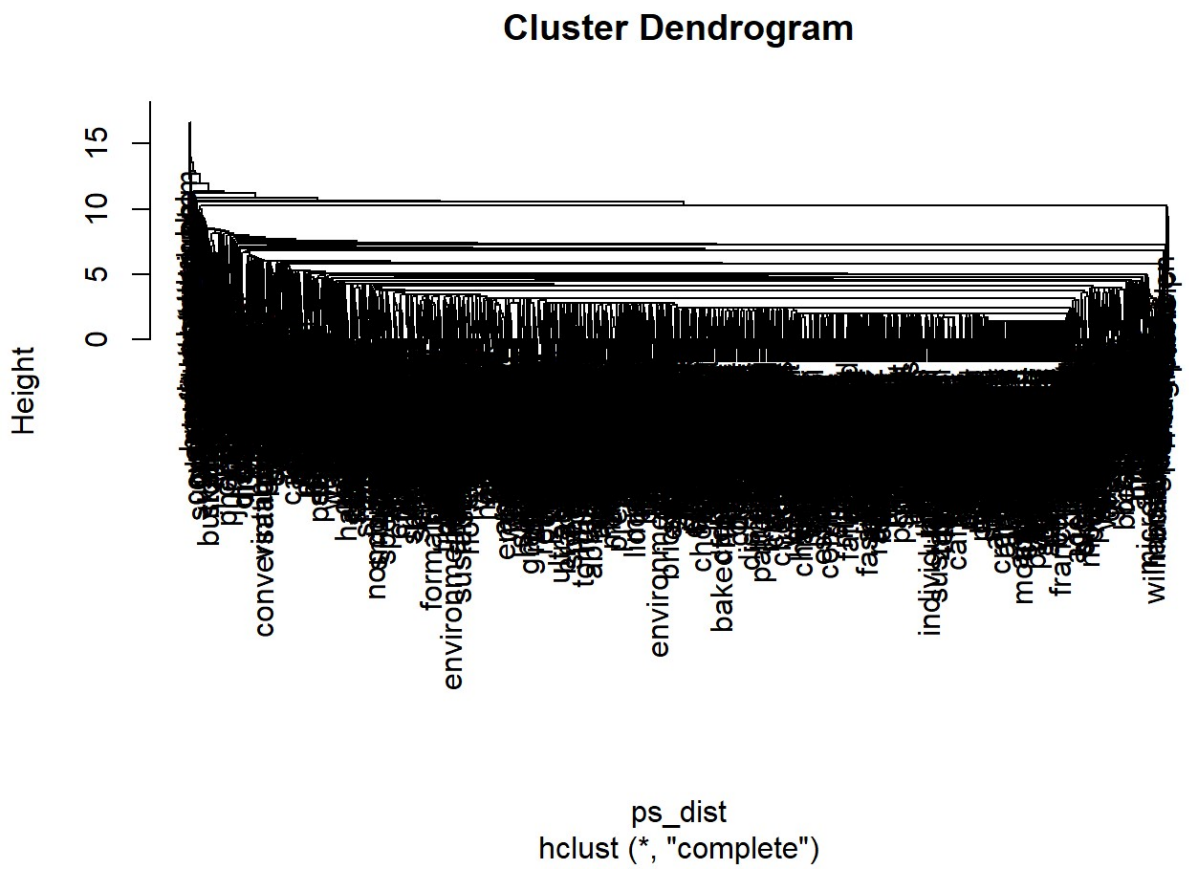
```
# Plot a barchart of the 25 most common words
ps_m<-as.matrix(TDM)
term_frequency<-rowSums(ps_m)
term_frequency<-sort(term_frequency, decreasing = TRUE)
barplot(term_frequency[1:25],col = "steel blue", las=2)
```




Above we have plotted 25 most common words

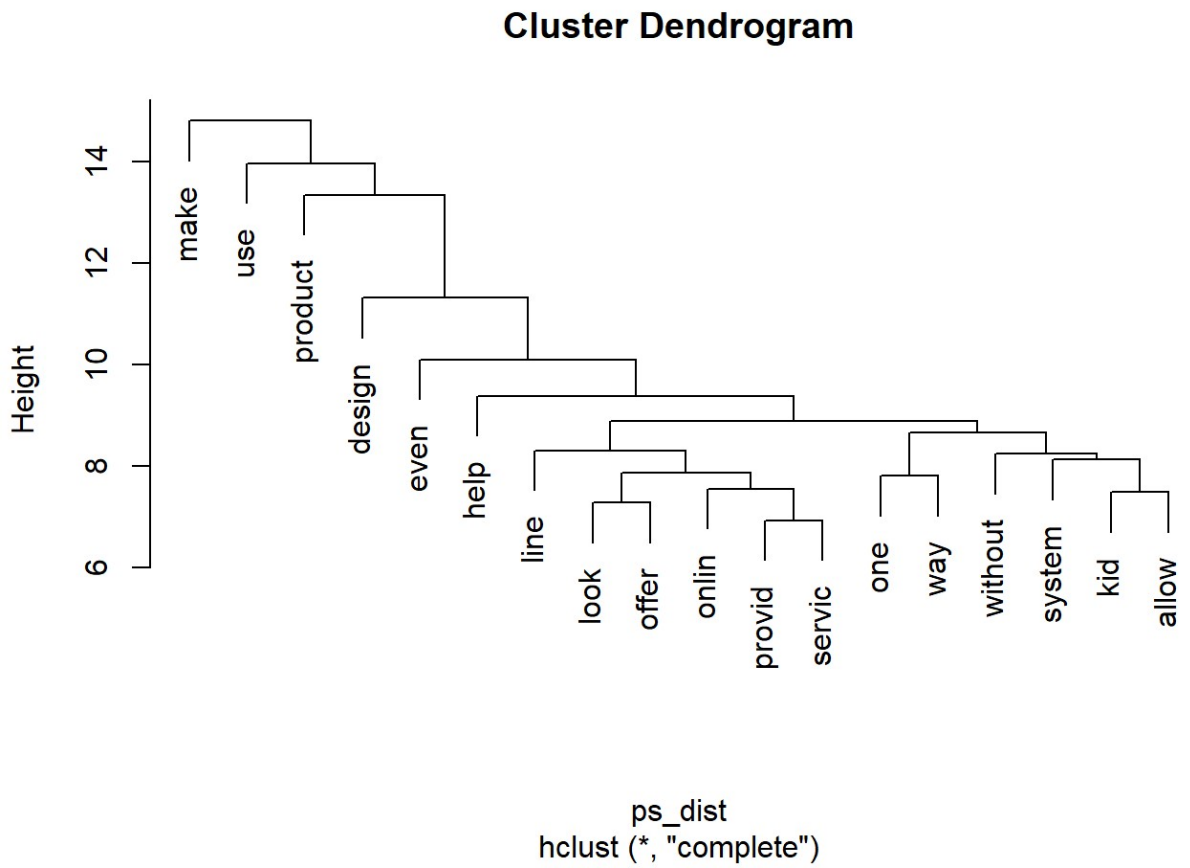
# Cluster analysis

```
#Cluster analysis
hc_df = as.data.frame(m1)
ps_dist = dist(hc_df, method = "euclidean")
hc = hclust(ps_dist)
plot(hc)
```




To reduce the dimensions in TDM, we will remove less frequent words using removeSparseTerms and sparsity less than 0.95

```
#Trim to study the top 50 terms
tdm1 = TDM[names(tail(sort(rowSums(as.matrix(TDM))),50)), ]
new_tdm<-removeSparseTerms(tdm1, sparse = 0.95)
tdm_m = as.matrix(new_tdm)
tdm_df = as.data.frame(tdm_m)
ps_dist = dist(tdm_df, method = "euclidean")
hc = hclust(ps_dist)
plot(hc)
```




```
dim(tdm_df)
```

```
## [1]  18 495
```

# Create associations

```
#Create associations_df#
#install.packages(qdap)
##library(qdap)##
#associations <- findAssocs(new_tdm,"make",0.05)
#associations_df <- list_vect2df(associations)[, 2:3]
#ggplot(associations_df, aes(y = associations_df[, 1])) +
#  geom_point(aes(x = associations_df[,2]),
#             data = associations_df, size = 3) +
#  ggtitle("Word Associations to 'make'") +
#  theme_gdocs()
```

```
#Use of N-grams
```

This the fnal data set prepareatio, here we convert this dataset into data.frame and add dependant

variable dealas final step

```
#Tokenize descriptions#
Sharktokens=tokens(Shark$description,what="word",
                   remove_numbers=TRUE,remove_punct=TRUE, remove_symbols=TRUE, remove
_hyphens=TRUE)
# Lowercase the tokens
Sharktokens=tokens_tolower(Sharktokens)
# remove stop words and unnecessary words
rmwords <- c( "etc","made_usa", "also", "xxs", "xs", "s","A","An","The","y")
Sharktokens=tokens_select(Sharktokens, stopwords(),selection = "remove")
Sharktokens=tokens_remove(Sharktokens,rmwords)
```

```
# Stemming tokens
Sharktokens=tokens_wordstem(Sharktokens,language = "english")
Sharktokens=tokens_ngrams(Sharktokens,n=1:2)
```

```
# Creating a bag of words #
Sharktokensdfm=dfm(Sharktokens,tolower = FALSE)
# Remove sparsity
SharkSparse <- convert(Sharktokensdfm, "tm")
tm::removeSparseTerms(SharkSparse, 0.99)
```

```
## <<DocumentTermMatrix (documents: 495, terms: 521)>>
## Non-/sparse entries: 5490/252405
## Sparsity           : 98%
## Maximal term length: 16
## Weighting          : term frequency (tf)
```

```
# Create the dfm
dfm_trim(Sharktokensdfm, min_docfreq = 0.2)
```

```
## Document-feature matrix of: 495 documents, 13,338 features (99.7% sparse).
```

```
x=dfm_trim(Sharktokensdfm, sparsity = 0.99)
```

```
## Setup a dataframe with features
df=convert(x,to="data.frame")
##Add the Y variable Recommend.IND
Sharktokensdf=cbind(Shark$deal,df)
names(Sharktokensdf)
```

```
##   [1] "Shark$deal"      "document"        "devic"
##   [4] "retail"          "two"             "new"
##   [7] "parent"          "young"           "children"
##  [10] "make"            "easi"            "littl"
##  [13] "one"             "turn"            "experi"
##  [16] "play"            "provid"          "posit"
##  [19] "organ"           "pack"            "move"
##  [22] "servic"          "deliv"           "colleg"
##  [25] "women"           "center"          "room"
##  [28] "offer"           "access"          "educ"
##  [31] "first"           "pitch"           "shark"
##  [34] "tank"            "perfect"         "line"
##  [37] "focus"           "gourmet"         "food"
##  [40] "product"         "sold"            "across"
##  [43] "store"           "includ"          "dress"
##  [46] "mani"            "design"          "flavor"
##  [49] "shark_tank"      "mix"             "cloth"
##  [52] "look"            "becom"           "big"
##  [55] "brand"           "activ"           "sport"
##  [58] "apparel"         "attach"          "hold"
##  [61] "side"            "screen"          "safeti"
##  [64] "prevent"         "start"           "vehicl"
##  [67] "hous"            "get"             "student"
##  [70] "learn"           "classic"         "work"
##  [73] "cook"            "household"       "item"
##  [76] "made"            "recycl"          "book"
##  [79] "help"            "find"            "fit"
##  [82] "fun"             "minut"           "effici"
##  [85] "entertain"       "us"              "set"
##  [88] "card"            "combin"          "keep"
##  [91] "kid"             "long"            "school"
##  [94] "develop"         "model"           "compani"
##  [97] "profession"      "remov"           "much"
## [100] "like"            "yet"             "babi"
## [103] "spend"           "time"            "can"
## [106] "rescu"           "size"            "quick"
## [109] "solut"           "add"             "anoth"
## [112] "layer"           "protect"         "surfac"
## [115] "machin"          "maintain"        "throw"
## [118] "wash"            "easier"          "tri"
## [121] "direct"          "bright"          "stylish"
## [124] "good"            "own"             "coffe"
## [127] "high"            "caffein"         "use"
## [130] "toy"             "base"            "buy"
## [133] "back"            "sell"            "gift"
## [136] "year"            "inspir"          "accessori"
## [139] "color"           "onlin"           "user"
## [142] "toward"          "well"            "bar"
```

```
## [145] "safe"             "enjoy"            "unlik"
## [148] "market"           "low"              "less"
## [151] "sugar"            "damag"            "ingredi"
## [154] "natur"            "test"             "ultim"
## [157] "just"             "band"             "weight"
## [160] "handl"            "famili"           "fashion"
## [163] "engin"            "bodi"             "order"
## [166] "energi"           "whole"            "custom"
## [169] "choos"            "go"               "person"
## [172] "made_order"       "natur_ingredi"    "beer"
## [175] "bottl"            "award"            "win"
## [178] "materi"           "around"           "leg"
## [181] "program"          "danc"             "live"
## [184] "attract"          "five"             "featur"
## [187] "stick"            "free"             "pair"
## [190] "three"            "instead"          "uniqu"
## [193] "belt"             "allow"            "stuf"
## [196] "anim"             "look_like"        "special"
## [199] "special_design"   "beverag"          "dog"
## [202] "golf"             "need"             "cover"
## [205] "run"              "replac"           "full"
## [208] "chain"            "blend"            "season"
## [211] "portabl"          "ball"             "patent"
## [214] "place"            "motorcycl"        "kind"
## [217] "eco"              "friend"           "eco_friend"
## [220] "individu"         "packag"           "clean"
## [223] "easili"           "bicycl"           "face"
## [226] "storag"           "pocket"           "shop"
## [229] "way"              "save"             "mobil"
## [232] "ice"              "cream"            "busi"
## [235] "ice_cream"        "ever"             "whether"
## [238] "regular"          "stand"            "print"
## [241] "see"              "sinc"             "take"
## [244] "still"            "even"             "show"
## [247] "manufactur"       "cost"             "age"
## [250] "shape"            "creat"            "encourag"
## [253] "pre"              "plan"             "usa"
## [256] "made_usa"         "hand"             "want"
## [259] "thing"            "practic"          "receiv"
## [262] "give"             "home"             "shoe"
## [265] "sale"             "parti"            "stay"
## [268] "without"          "struggl"          "plastic"
## [271] "men"              "signatur"         "power"
## [274] "secur"            "strong"           "magnet"
## [277] "shirt"            "produc"           "expand"
## [280] "suit"             "money"            "clip"
## [283] "tie"              "travel"           "bag"
## [286] "premium"          "wine"             "glass"
## [289] "cup"              "flagship"         "origin"
```

```
## [292] "flagship_product" "subscript"        "fresh"
## [295] "bring"            "self"             "air"
## [298] "come"             "reduc"            "effect"
## [301] "super"            "lightweight"      "ultra"
## [304] "durabl"           "ride"             "tradit"
## [307] "simpli"           "bike"             "feet"
## [310] "fulli"            "readi"            "resist"
## [313] "mom"              "qualiti"          "care"
## [316] "craft"            "piec"             "serv"
## [319] "found"            "current"          "expens"
## [322] "sourc"            "sustain"          "wood"
## [325] "high_qualiti"     "avail"            "choic"
## [328] "let"              "kid_can"          "anyon"
## [331] "pet"              "smell"            "differ"
## [334] "spray"            "chemic"           "continu"
## [337] "can_buy"          "countri"          "style"
## [340] "truck"            "stop"             "seat"
## [343] "addit"            "preserv"          "lip"
## [346] "light"            "car"              "window"
## [349] "small"            "case"             "gear"
## [352] "skin"             "appli"            "built"
## [355] "cake"             "recip"            "ship"
## [358] "system"           "pad"              "train"
## [361] "video"            "game"             "altern"
## [364] "cours"            "among"            "non"
## [367] "wed"              "event"            "fast"
## [370] "fire"             "connect"          "tool"
## [373] "kit"              "box"              "success"
## [376] "healthi"          "month"            "display"
## [379] "interchang"       "allow_user"       "theme"
## [382] "make_fashion"     "water"            "contain"
## [385] "top"              "eas"              "websit"
## [388] "peopl"            "purchas"          "partner"
## [391] "larg"             "prospect"         "transform"
## [394] "carri"            "bed"              "magic"
## [397] "day"              "electr"           "right"
## [400] "music"            "real"             "favorit"
## [403] "instruct"         "open"             "point"
## [406] "serious"          "butter"           "oil"
## [409] "gluten"           "now"              "name"
## [412] "insid"            "inflat"           "chang"
## [415] "better"           "comfort"          "adjust"
## [418] "varieti"          "best"             "space"
## [421] "flat"             "creativ"          "chocol"
## [424] "corpor"           "improv"           "equip"
## [427] "technolog"        "although"         "number"
## [430] "might"            "conveni"          "rest"
## [433] "alway"            "rather"           "smart"
## [436] "kitchen"          "treat"            "cut"
```

```
## [439] "paper"           "accord"           "can_use"
## [442] "price"           "togeth"           "everi"
## [445] "rang"            "premium_qualiti"  "meal"
## [448] "portion"         "wast"             "workout"
## [451] "secret"          "visibl"           "wearer"
## [454] "mother"          "drink"            "put"
## [457] "thank"           "wheel"            "inch"
## [460] "nut"             "singl"            "peanut"
## [463] "social"          "call"             "smartphon"
## [466] "build"           "app"              "reusabl"
## [469] "sturdi"          "phone"            "grow"
## [472] "afford"          "content"          "consum"
## [475] "pay"             "may"              "feel"
## [478] "form"            "ad"               "roll"
## [481] "california"      "balanc"           "control"
## [484] "support"         "cool"             "heat"
## [487] "function"        "remot"            "tast"
## [490] "hair"            "outdoor"          "innov"
## [493] "invent"          "love"             "sound"
## [496] "snack"           "someth"           "result"
## [499] "process"         "delici"           "versatil"
## [502] "goe"             "simpl"            "complet"
## [505] "mission"         "detail"           "old"
## [508] "great"           "audio"            "promot"
## [511] "iphon"           "nation"           "team"
## [514] "duo"             "behind"           "tablet"
## [517] "meet"            "select"           "world"
## [520] "respons"         "beach"            "typic"
## [523] "companion"
```

# Final data set

```
## Cleanup names
names(Sharktokensdf)[names(Sharktokensdf) == "Shark$deal"] <- "deal"
#names(Sharktokensdf)=make.names(names(Sharktokensdf))
head(Sharktokensdf)
```

```
##    deal document devic retail two new parent young children make easi
## 1 FALSE    text1     1      0   0   0      0     0        0    0    0
## 2  TRUE    text2     0      2   1   1      0     0        0    0    0
## 3  TRUE    text3     0      0   0   0      1     1        1    1    1
## 4 FALSE    text4     0      0   0   0      0     0        0    0    0
## 5 FALSE    text5     0      0   0   0      0     0        0    0    0
## 6  TRUE    text6     0      1   0   0      0     0        0    0    0
##   littl one turn experi play provid posit organ pack move servic deliv
## 1     0   0    0      0    0      0     0     0    0    0      0     0
## 2     0   0    0      0    0      0     0     0    0    0      0     0
## 3     1   1    1      1    1      1     1     0    0    0      0     0
## 4     0   0    0      0    0      0     0     1    1    1      1     1
## 5     0   0    0      0    0      0     0     0    0    0      0     0
## 6     0   1    0      0    0      0     0     0    0    0      0     0
##   colleg women center room offer access educ first pitch shark tank
## 1      0     0      0    0     0      0    0     0     0     0    0
## 2      0     0      0    0     0      0    0     0     0     0    0
## 3      0     0      0    0     0      0    0     0     0     0    0
## 4      1     1      0    0     0      0    0     0     0     0    0
## 5      0     0      1    1     1      1    1     0     0     0    0
## 6      0     0      0    0     0      0    0     1     1     1    1
##   perfect line focus gourmet food product sold across store includ dress
## 1       0    0     0       0    0       0    0      0     0      0     0
## 2       0    0     0       0    0       0    0      0     0      0     0
## 3       0    0     0       0    0       0    0      0     0      0     0
## 4       0    0     0       0    0       0    0      0     0      0     0
## 5       0    0     0       0    0       0    0      0     0      0     0
## 6       2    1     1       1    1       2    1      1     1      1     1
##   mani design flavor shark_tank mix cloth look becom big brand activ sport
## 1    0      0      0          0   0     0    0     0   0     0     0     0
## 2    0      0      0          0   0     0    0     0   0     0     0     0
## 3    0      0      0          0   0     0    0     0   0     0     0     0
## 4    0      0      0          0   0     0    0     0   0     0     0     0
## 5    0      0      0          0   0     0    0     0   0     0     0     0
## 6    1      1      1          1   0     0    0     0   0     0     0     0
##   apparel attach hold side screen safeti prevent start vehicl hous get
## 1       0      0    0    0      0      0       0     0      0    0   0
## 2       0      0    0    0      0      0       0     0      0    0   0
## 3       0      0    0    0      0      0       0     0      0    0   0
## 4       0      0    0    0      0      0       0     0      0    0   0
## 5       0      0    0    0      0      0       0     0      0    0   0
## 6       0      0    0    0      0      0       0     0      0    0   0
##   student learn classic work cook household item made recycl book help
## 1       0     0       0    0    0         0    0    0      0    0    0
## 2       0     0       0    0    0         0    0    0      0    0    0
## 3       0     0       0    0    0         0    0    0      0    0    0
## 4       0     0       0    0    0         0    0    0      0    0    0
## 5       0     0       0    0    0         0    0    0      0    0    0
```

```
## 6       0     0       0    0    0         0    0    0      0    0    0
##   find fit fun minut effici entertain us set card combin keep kid long
## 1    0   0   0     0      0         0  0   0    0      0    0   0    0
## 2    0   0   0     0      0         0  0   0    0      0    0   0    0
## 3    0   0   0     0      0         0  0   0    0      0    0   0    0
## 4    0   0   0     0      0         0  0   0    0      0    0   0    0
## 5    0   0   0     0      0         0  0   0    0      0    0   0    0
## 6    0   0   0     0      0         0  0   0    0      0    0   0    0
##   school develop model compani profession remov much like yet babi spend
## 1      0       0     0       0          0     0    0    0   0    0     0
## 2      0       0     0       0          0     0    0    0   0    0     0
## 3      0       0     0       0          0     0    0    0   0    0     0
## 4      0       0     0       0          0     0    0    0   0    0     0
## 5      0       0     0       0          0     0    0    0   0    0     0
## 6      0       0     0       0          0     0    0    0   0    0     0
##   time can rescu size quick solut add anoth layer protect surfac machin
## 1    0   0     0    0     0     0   0     0     0       0      0      0
## 2    0   0     0    0     0     0   0     0     0       0      0      0
## 3    0   0     0    0     0     0   0     0     0       0      0      0
## 4    0   0     0    0     0     0   0     0     0       0      0      0
## 5    0   0     0    0     0     0   0     0     0       0      0      0
## 6    0   0     0    0     0     0   0     0     0       0      0      0
##   maintain throw wash easier tri direct bright stylish good own coffe high
## 1        0     0    0      0   0      0      0       0    0   0     0    0
## 2        0     0    0      0   0      0      0       0    0   0     0    0
## 3        0     0    0      0   0      0      0       0    0   0     0    0
## 4        0     0    0      0   0      0      0       0    0   0     0    0
## 5        0     0    0      0   0      0      0       0    0   0     0    0
## 6        0     0    0      0   0      0      0       0    0   0     0    0
##   caffein use toy base buy back sell gift year inspir accessori color
## 1       0   0   0    0   0    0    0    0    0      0         0     0
## 2       0   0   0    0   0    0    0    0    0      0         0     0
## 3       0   0   0    0   0    0    0    0    0      0         0     0
## 4       0   0   0    0   0    0    0    0    0      0         0     0
## 5       0   0   0    0   0    0    0    0    0      0         0     0
## 6       0   0   0    0   0    0    0    0    0      0         0     0
##   onlin user toward well bar safe enjoy unlik market low less sugar damag
## 1     0    0      0    0   0    0     0     0      0   0    0     0     0
## 2     0    0      0    0   0    0     0     0      0   0    0     0     0
## 3     0    0      0    0   0    0     0     0      0   0    0     0     0
## 4     0    0      0    0   0    0     0     0      0   0    0     0     0
## 5     0    0      0    0   0    0     0     0      0   0    0     0     0
## 6     0    0      0    0   0    0     0     0      0   0    0     0     0
##   ingredi natur test ultim just band weight handl famili fashion engin
## 1       0     0    0     0    0    0      0     0      0       0     0
## 2       0     0    0     0    0    0      0     0      0       0     0
## 3       0     0    0     0    0    0      0     0      0       0     0
## 4       0     0    0     0    0    0      0     0      0       0     0
## 5       0     0    0     0    0    0      0     0      0       0     0
```

```
## 6       0     0    0     0    0    0      0     0      0       0     0
##   bodi order energi whole custom choos go person made_order natur_ingredi
## 1    0     0      0     0      0     0  0      0          0             0
## 2    0     0      0     0      0     0  0      0          0             0
## 3    0     0      0     0      0     0  0      0          0             0
## 4    0     0      0     0      0     0  0      0          0             0
## 5    0     0      0     0      0     0  0      0          0             0
## 6    0     0      0     0      0     0  0      0          0             0
##   beer bottl award win materi around leg program danc live attract five
## 1    0     0     0   0      0      0   0       0    0    0       0    0
## 2    0     0     0   0      0      0   0       0    0    0       0    0
## 3    0     0     0   0      0      0   0       0    0    0       0    0
## 4    0     0     0   0      0      0   0       0    0    0       0    0
## 5    0     0     0   0      0      0   0       0    0    0       0    0
## 6    0     0     0   0      0      0   0       0    0    0       0    0
##   featur stick free pair three instead uniqu belt allow stuf anim
## 1      0     0    0    0     0       0     0    0     0    0    0
## 2      0     0    0    0     0       0     0    0     0    0    0
## 3      0     0    0    0     0       0     0    0     0    0    0
## 4      0     0    0    0     0       0     0    0     0    0    0
## 5      0     0    0    0     0       0     0    0     0    0    0
## 6      0     0    0    0     0       0     0    0     0    0    0
##   look_like special special_design beverag dog golf need cover run replac
## 1         0       0              0       0   0    0    0     0   0      0
## 2         0       0              0       0   0    0    0     0   0      0
## 3         0       0              0       0   0    0    0     0   0      0
## 4         0       0              0       0   0    0    0     0   0      0
## 5         0       0              0       0   0    0    0     0   0      0
## 6         0       0              0       0   0    0    0     0   0      0
##   full chain blend season portabl ball patent place motorcycl kind eco
## 1    0     0     0      0       0    0      0     0         0    0   0
## 2    0     0     0      0       0    0      0     0         0    0   0
## 3    0     0     0      0       0    0      0     0         0    0   0
## 4    0     0     0      0       0    0      0     0         0    0   0
## 5    0     0     0      0       0    0      0     0         0    0   0
## 6    0     0     0      0       0    0      0     0         0    0   0
##   friend eco_friend individu packag clean easili bicycl face storag pocket
## 1      0          0        0      0     0      0      0    0      0      0
## 2      0          0        0      0     0      0      0    0      0      0
## 3      0          0        0      0     0      0      0    0      0      0
## 4      0          0        0      0     0      0      0    0      0      0
## 5      0          0        0      0     0      0      0    0      0      0
## 6      0          0        0      0     0      0      0    0      0      0
##   shop way save mobil ice cream busi ice_cream ever whether regular stand
## 1    0   0    0     0   0     0    0         0    0       0       0     0
## 2    0   0    0     0   0     0    0         0    0       0       0     0
## 3    0   0    0     0   0     0    0         0    0       0       0     0
## 4    0   0    0     0   0     0    0         0    0       0       0     0
## 5    0   0    0     0   0     0    0         0    0       0       0     0
```

```
## 6    0   0    0     0   0     0    0         0    0       0       0     0
##   print see sinc take still even show manufactur cost age shape creat
## 1     0   0    0    0     0    0    0          0    0   0     0     0
## 2     0   0    0    0     0    0    0          0    0   0     0     0
## 3     0   0    0    0     0    0    0          0    0   0     0     0
## 4     0   0    0    0     0    0    0          0    0   0     0     0
## 5     0   0    0    0     0    0    0          0    0   0     0     0
## 6     0   0    0    0     0    0    0          0    0   0     0     0
##   encourag pre plan usa made_usa hand want thing practic receiv give home
## 1        0   0    0   0        0    0    0     0       0      0    0    0
## 2        0   0    0   0        0    0    0     0       0      0    0    0
## 3        0   0    0   0        0    0    0     0       0      0    0    0
## 4        0   0    0   0        0    0    0     0       0      0    0    0
## 5        0   0    0   0        0    0    0     0       0      0    0    0
## 6        0   0    0   0        0    0    0     0       0      0    0    0
##   shoe sale parti stay without struggl plastic men signatur power secur
## 1    0    0     0    0       0       0       0   0        0     0     0
## 2    0    0     0    0       0       0       0   0        0     0     0
## 3    0    0     0    0       0       0       0   0        0     0     0
## 4    0    0     0    0       0       0       0   0        0     0     0
## 5    0    0     0    0       0       0       0   0        0     0     0
## 6    0    0     0    0       0       0       0   0        0     0     0
##   strong magnet shirt produc expand suit money clip tie travel bag premium
## 1      0      0     0      0      0    0     0    0   0      0   0       0
## 2      0      0     0      0      0    0     0    0   0      0   0       0
## 3      0      0     0      0      0    0     0    0   0      0   0       0
## 4      0      0     0      0      0    0     0    0   0      0   0       0
## 5      0      0     0      0      0    0     0    0   0      0   0       0
## 6      0      0     0      0      0    0     0    0   0      0   0       0
##   wine glass cup flagship origin flagship_product subscript fresh bring
## 1    0     0   0        0      0                0         0     0     0
## 2    0     0   0        0      0                0         0     0     0
## 3    0     0   0        0      0                0         0     0     0
## 4    0     0   0        0      0                0         0     0     0
## 5    0     0   0        0      0                0         0     0     0
## 6    0     0   0        0      0                0         0     0     0
##   self air come reduc effect super lightweight ultra durabl ride tradit
## 1    0   0    0     0      0     0           0     0      0    0      0
## 2    0   0    0     0      0     0           0     0      0    0      0
## 3    0   0    0     0      0     0           0     0      0    0      0
## 4    0   0    0     0      0     0           0     0      0    0      0
## 5    0   0    0     0      0     0           0     0      0    0      0
## 6    0   0    0     0      0     0           0     0      0    0      0
##   simpli bike feet fulli readi resist mom qualiti care craft piec serv
## 1      0    0    0     0     0      0   0       0    0     0    0    0
## 2      0    0    0     0     0      0   0       0    0     0    0    0
## 3      0    0    0     0     0      0   0       0    0     0    0    0
## 4      0    0    0     0     0      0   0       0    0     0    0    0
## 5      0    0    0     0     0      0   0       0    0     0    0    0
```

```
## 6      0    0    0     0     0      0   0       0    0     0    0    0
##   found current expens sourc sustain wood high_qualiti avail choic let
## 1     0       0      0     0       0    0            0     0     0   0
## 2     0       0      0     0       0    0            0     0     0   0
## 3     0       0      0     0       0    0            0     0     0   0
## 4     0       0      0     0       0    0            0     0     0   0
## 5     0       0      0     0       0    0            0     0     0   0
## 6     0       0      0     0       0    0            0     0     0   0
##   kid_can anyon pet smell differ spray chemic continu can_buy countri
## 1       0     0   0     0      0     0      0       0       0       0
## 2       0     0   0     0      0     0      0       0       0       0
## 3       0     0   0     0      0     0      0       0       0       0
## 4       0     0   0     0      0     0      0       0       0       0
## 5       0     0   0     0      0     0      0       0       0       0
## 6       0     0   0     0      0     0      0       0       0       0
##   style truck stop seat addit preserv lip light car window small case gear
## 1     0     0    0    0     0       0   0     0   0      0     0    0    0
## 2     0     0    0    0     0       0   0     0   0      0     0    0    0
## 3     0     0    0    0     0       0   0     0   0      0     0    0    0
## 4     0     0    0    0     0       0   0     0   0      0     0    0    0
## 5     0     0    0    0     0       0   0     0   0      0     0    0    0
## 6     0     0    0    0     0       0   0     0   0      0     0    0    0
##   skin appli built cake recip ship system pad train video game altern
## 1    0     0     0    0     0    0      0   0     0     0    0       0
## 2    0     0     0    0     0    0      0   0     0     0    0       0
## 3    0     0     0    0     0    0      0   0     0     0    0       0
## 4    0     0     0    0     0    0      0   0     0     0    0       0
## 5    0     0     0    0     0    0      0   0     0     0    0       0
## 6    0     0     0    0     0    0      0   0     0     0    0       0
##   cours among non wed event fast fire connect tool kit box success healthi
## 1     0     0   0   0     0    0    0       0    0   0   0       0       0
## 2     0     0   0   0     0    0    0       0    0   0   0       0       0
## 3     0     0   0   0     0    0    0       0    0   0   0       0       0
## 4     0     0   0   0     0    0    0       0    0   0   0       0       0
## 5     0     0   0   0     0    0    0       0    0   0   0       0       0
## 6     0     0   0   0     0    0    0       0    0   0   0       0       0
##   month display interchang allow_user theme make_fashion water contain top
## 1     0       0          0          0     0            0     0       0   0
## 2     0       0          0          0     0            0     0       0   0
## 3     0       0          0          0     0            0     0       0   0
## 4     0       0          0          0     0            0     0       0   0
## 5     0       0          0          0     0            0     0       0   0
## 6     0       0          0          0     0            0     0       0   0
##   eas websit peopl purchas partner larg prospect transform carri bed magic
## 1   0      0     0       0       0    0        0         0     0   0     0
## 2   0      0     0       0       0    0        0         0     0   0     0
## 3   0      0     0       0       0    0        0         0     0   0     0
## 4   0      0     0       0       0    0        0         0     0   0     0
## 5   0      0     0       0       0    0        0         0     0   0     0
```

```
## 6   0      0     0       0       0    0        0         0     0   0     0
##   day electr right music real favorit instruct open point serious butter
## 1   0      0     0     0    0       0        0    0     0       0      0
## 2   0      0     0     0    0       0        0    0     0       0      0
## 3   0      0     0     0    0       0        0    0     0       0      0
## 4   0      0     0     0    0       0        0    0     0       0      0
## 5   0      0     0     0    0       0        0    0     0       0      0
## 6   0      0     0     0    0       0        0    0     0       0      0
##   oil gluten now name insid inflat chang better comfort adjust varieti
## 1   0      0   0    0     0      0     0      0       0      0       0
## 2   0      0   0    0     0      0     0      0       0      0       0
## 3   0      0   0    0     0      0     0      0       0      0       0
## 4   0      0   0    0     0      0     0      0       0      0       0
## 5   0      0   0    0     0      0     0      0       0      0       0
## 6   0      0   0    0     0      0     0      0       0      0       0
##   best space flat creativ chocol corpor improv equip technolog although
## 1    0     0    0       0      0      0      0     0         0        0
## 2    0     0    0       0      0      0      0     0         0        0
## 3    0     0    0       0      0      0      0     0         0        0
## 4    0     0    0       0      0      0      0     0         0        0
## 5    0     0    0       0      0      0      0     0         0        0
## 6    0     0    0       0      0      0      0     0         0        0
##   number might conveni rest alway rather smart kitchen treat cut paper
## 1      0     0       0    0     0      0     0       0     0   0     0
## 2      0     0       0    0     0      0     0       0     0   0     0
## 3      0     0       0    0     0      0     0       0     0   0     0
## 4      0     0       0    0     0      0     0       0     0   0     0
## 5      0     0       0    0     0      0     0       0     0   0     0
## 6      0     0       0    0     0      0     0       0     0   0     0
##   accord can_use price togeth everi rang premium_qualiti meal portion wast
## 1      0       0     0      0     0    0               0    0       0    0
## 2      0       0     0      0     0    0               0    0       0    0
## 3      0       0     0      0     0    0               0    0       0    0
## 4      0       0     0      0     0    0               0    0       0    0
## 5      0       0     0      0     0    0               0    0       0    0
## 6      0       0     0      0     0    0               0    0       0    0
##   workout secret visibl wearer mother drink put thank wheel inch nut singl
## 1       0      0      0      0      0     0   0     0     0    0   0     0
## 2       0      0      0      0      0     0   0     0     0    0   0     0
## 3       0      0      0      0      0     0   0     0     0    0   0     0
## 4       0      0      0      0      0     0   0     0     0    0   0     0
## 5       0      0      0      0      0     0   0     0     0    0   0     0
## 6       0      0      0      0      0     0   0     0     0    0   0     0
##   peanut social call smartphon build app reusabl sturdi phone grow afford
## 1      0      0    0         0     0   0       0      0     0    0      0
## 2      0      0    0         0     0   0       0      0     0    0      0
## 3      0      0    0         0     0   0       0      0     0    0      0
## 4      0      0    0         0     0   0       0      0     0    0      0
## 5      0      0    0         0     0   0       0      0     0    0      0
```

```
## 6      0      0    0         0     0   0       0      0     0    0      0
##   content consum pay may feel form ad roll california balanc control
## 1       0      0   0   0    0    0  0    0          0      0       0
## 2       0      0   0   0    0    0  0    0          0      0       0
## 3       0      0   0   0    0    0  0    0          0      0       0
## 4       0      0   0   0    0    0  0    0          0      0       0
## 5       0      0   0   0    0    0  0    0          0      0       0
## 6       0      0   0   0    0    0  0    0          0      0       0
##   support cool heat function remot tast hair outdoor innov invent love
## 1       0    0    0        0     0    0    0       0     0      0    0
## 2       0    0    0        0     0    0    0       0     0      0    0
## 3       0    0    0        0     0    0    0       0     0      0    0
## 4       0    0    0        0     0    0    0       0     0      0    0
## 5       0    0    0        0     0    0    0       0     0      0    0
## 6       0    0    0        0     0    0    0       0     0      0    0
##   sound snack someth result process delici versatil goe simpl complet
## 1     0     0      0      0       0      0        0   0     0       0
## 2     0     0      0      0       0      0        0   0     0       0
## 3     0     0      0      0       0      0        0   0     0       0
## 4     0     0      0      0       0      0        0   0     0       0
## 5     0     0      0      0       0      0        0   0     0       0
## 6     0     0      0      0       0      0        0   0     0       0
##   mission detail old great audio promot iphon nation team duo behind
## 1       0      0   0     0     0      0     0      0    0   0      0
## 2       0      0   0     0     0      0     0      0    0   0      0
## 3       0      0   0     0     0      0     0      0    0   0      0
## 4       0      0   0     0     0      0     0      0    0   0      0
## 5       0      0   0     0     0      0     0      0    0   0      0
## 6       0      0   0     0     0      0     0      0    0   0      0
##   tablet meet select world respons beach typic companion
## 1      0    0      0     0       0     0     0         0
## 2      0    0      0     0       0     0     0         0
## 3      0    0      0     0       0     0     0         0
## 4      0    0      0     0       0     0     0         0
## 5      0    0      0     0       0     0     0         0
## 6      0    0      0     0       0     0     0         0
```

```
## Remove the original review.text column
data_shark=Sharktokensdf[,-c(2)]
data_shark$deal<-ifelse(data_shark$deal=="TRUE",1,0)
```

# Data Preparation

```
str(data_shark)
```

```
## 'data.frame':    495 obs. of  522 variables:
##  $ deal            : num  0 1 1 0 0 1 0 0 0 1 ...
##  $ devic           : num  1 0 0 0 0 0 0 0 1 0 ...
##  $ retail          : num  0 2 0 0 0 1 0 0 0 0 ...
##  $ two             : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ new             : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ parent          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ young           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ children        : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ make            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ easi            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ littl           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ one             : num  0 0 1 0 0 1 0 0 0 0 ...
##  $ turn            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ experi          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ play            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ provid          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ posit           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ organ           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ pack            : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ move            : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ servic          : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ deliv           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ colleg          : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ women           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ center          : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ room            : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ offer           : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ access          : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ educ            : num  0 0 0 0 1 0 0 0 0 1 ...
##  $ first           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ pitch           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ shark           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ tank            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ perfect         : num  0 0 0 0 0 2 0 0 0 0 ...
##  $ line            : num  0 0 0 0 0 1 1 0 0 0 ...
##  $ focus           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ gourmet         : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ food            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ product         : num  0 0 0 0 0 2 0 0 0 0 ...
##  $ sold            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ across          : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ store           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ includ          : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ dress           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ mani            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ design          : num  0 0 0 0 0 1 0 0 0 1 ...
##  $ flavor          : num  0 0 0 0 0 1 0 0 0 0 ...
```

```
##  $ shark_tank      : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ mix             : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ cloth           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ look            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ becom           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ big             : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ brand           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ activ           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ sport           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ apparel         : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ attach          : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ hold            : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ side            : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ screen          : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ safeti          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ prevent         : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ start           : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ vehicl          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ hous            : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ get             : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ student         : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ learn           : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ classic         : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ work            : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ cook            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ household       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ item            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ made            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ recycl          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ book            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ help            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ find            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ fit             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ fun             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ minut           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ effici          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ entertain       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ us              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ set             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ card            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ combin          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ keep            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ kid             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ long            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ school          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ develop         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ model           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ compani         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ profession      : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ remov           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ much            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ like            : num  0 0 0 0 0 0 0 0 0 0 ...
##   [list output truncated]
```

We observed that the final data set is contains 495 observation with 522 variables:

## Check the proportion of data

```
#Check the proportion of data
nrow(subset(data_shark, deal == 1))/nrow(data_shark)
```

```
## [1] 0.5070707
```

```
set.seed(123)
train_idx <- sample(c(1:nrow(data_shark)), round(nrow(data_shark) * 0.7,0), replace =
FALSE)
train_data <- data_shark[train_idx,]
test_data <- data_shark[-train_idx,]
dim(train_data)
```

```
## [1] 346 522
```

```
dim(test_data)
```

```
## [1] 149 522
```

Observed that there are 50% of True values whereas the traine data consist of 346/522, and testdata set is 149/522. Further we are going to making the sampling balcne before preparing the modesls

```
train.pos <- subset(train_data, deal == 1)
train.neg <- subset(train_data, deal == 0)
dim(train.pos)
```

```
## [1] 173 522
```

```
dim(train.neg)
```

```
## [1] 173 522
```

```
## Set the seed
set.seed(108)
## Take the sample subset from the major class (here negative)
train.neg.sub_idx <- sample(c(1:nrow(train.neg)), nrow(train.pos), replace = FALSE)
train_new <- train.neg[train.neg.sub_idx,]
dim(train_new)
```

```
## [1] 173 522
```

# Merge the negative and positive cases

```
#Merge the negative and positive cases
train_new <- rbind(train_new, train.pos)
dim(train_new)
```

```
## [1] 346 522
```

```
#Rendomizing  the data
train_new <- train_new[sample(1:nrow(train_new)),]
```

# Checking the proportion of deal in the sample

```
## Now check the proportion of deal in the sample
## in train_data
nrow(subset(train_data, deal == 1))/nrow(train_data)
```

```
## [1] 0.5
```

```
## in train.new
nrow(subset(train_new, deal == 1))/nrow(train_new)
```

```
## [1] 0.5
```

```
str(train_new)
```

```
## 'data.frame':    346 obs. of  522 variables:
##  $ deal            : num  1 0 0 0 0 0 1 1 0 0 ...
##  $ devic           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ retail          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ two             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ new             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ parent          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ young           : num  0 0 0 0 0 0 0 2 0 0 ...
##  $ children        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ make            : num  0 0 0 0 0 1 2 1 0 1 ...
##  $ easi            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ littl           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ one             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ turn            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ experi          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ play            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ provid          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ posit           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ organ           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pack            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ move            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ servic          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ deliv           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ colleg          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ women           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ center          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ room            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ offer           : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ access          : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ educ            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ first           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ pitch           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ shark           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ tank            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ perfect         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ line            : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ focus           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ gourmet         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ food            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ product         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sold            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ across          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ store           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ includ          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ dress           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ mani            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ design          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ flavor          : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ shark_tank      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ mix             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ cloth           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ look            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ becom           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ big             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ brand           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ activ           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sport           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ apparel         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ attach          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ hold            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ side            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ screen          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ safeti          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ prevent         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ start           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ vehicl          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ hous            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ get             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ student         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ learn           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ classic         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ work            : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ cook            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ household       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ item            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ made            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ recycl          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ book            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ help            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ find            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ fit             : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ fun             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ minut           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ effici          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ entertain       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ us              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ set             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ card            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ combin          : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ keep            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ kid             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ long            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ school          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ develop         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ model           : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ compani         : num  0 0 0 1 0 0 0 1 0 0 ...
##  $ profession      : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ remov           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ much            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ like            : num  0 0 0 0 0 0 0 0 0 0 ...
##   [list output truncated]
```

```
test_new <- (test_data)
```

# Model Planning and Building

We are planning for three models as follows: 1. CART 2. Logistic Regression 3. RandomForst To predict whether investors(aka shark) will invest in the businesses we will use deal as an output variable and use the CART, logistic regression and random forest models to measure the performance and accuracy of the model.

# CART Model

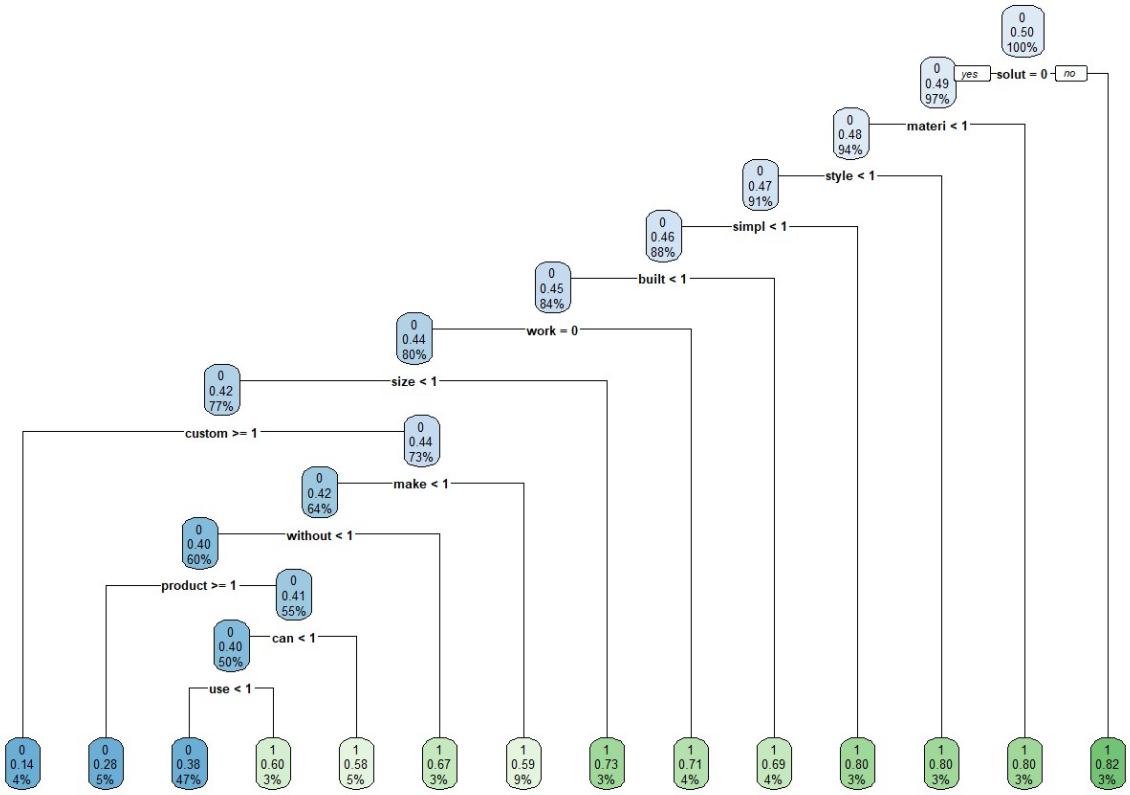
## Evaluate the performance of the CART model

```
library(rpart)
```

```
##
## Attaching package: 'rpart'
```

```
## The following object is masked from 'package:dendextend':
##
##     prune
```

```
library(rpart.plot)
r.ctrl = rpart.control(minsplit = 100, minbucket = 10, cp = 0, xval = 10)
DTmodel = rpart(deal ~., data = train_new, method = "class", control = r.ctrl)
rpart.plot(DTmodel)
```


0
0.50
100%
0
0.49
97%
yes
solut = 0
no
0
0.48
94%
materi < 1
0
0.47
91%
style < 1
0
0.46
88%
simpl < 1
0
0.45
84%
built < 1
0
0.44
80%
work = 0
0
0.42
77%
size < 1
custom >= 1
0
0.44
73%
0
0.42
64%
make < 1
0
0.40
60%
without < 1
product >= 1
0
0.41
55%
0
0.40
50%
can < 1
use < 1
0
0.14
4%
0
0.28
5%
0
0.38
47%
1
0.60
3%
1
0.58
5%
1
0.67
3%
1
0.59
9%
1
0.73
3%
1
0.71
4%
1
0.69
4%
1
0.80
3%
1
0.80
3%
1
0.80
3%
1
0.82
3%

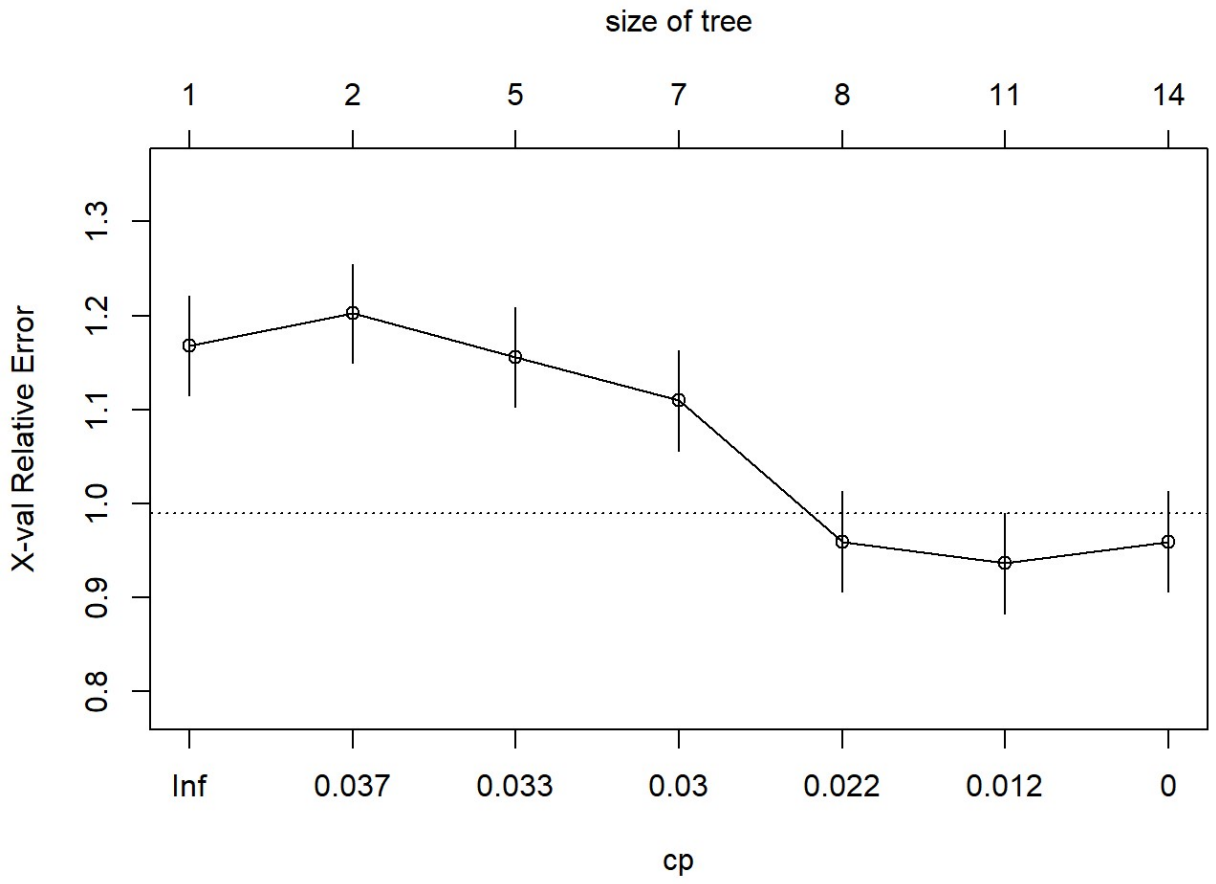
```
DTmodel
```

```
## n= 346
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##    1) root 346 173 0 (0.5000000 0.5000000)
##      2) solut< 0.5 335 164 0 (0.5104478 0.4895522)
##        4) materi< 0.5 325 156 0 (0.5200000 0.4800000)
##          8) style< 0.5 315 148 0 (0.5301587 0.4698413)
##           16) simpl< 0.5 305 140 0 (0.5409836 0.4590164)
##             32) built< 0.5 292 131 0 (0.5513699 0.4486301)
##               64) work< 0.5 278 121 0 (0.5647482 0.4352518)
##                128) size< 0.5 267 113 0 (0.5767790 0.4232210)
##                  256) custom>=0.5 14   2 0 (0.8571429 0.1428571) *
##                  257) custom< 0.5 253 111 0 (0.5612648 0.4387352)
##                    514) make< 0.5 221  92 0 (0.5837104 0.4162896)
##                     1028) without< 0.5 209  84 0 (0.5980861 0.4019139)
##                       2056) product>=0.5 18   5 0 (0.7222222 0.2777778) *
##                       2057) product< 0.5 191  79 0 (0.5863874 0.4136126)
##                         4114) can< 0.5 172  68 0 (0.6046512 0.3953488)
##                           8228) use< 0.5 162  62 0 (0.6172840 0.3827160) *
##                           8229) use>=0.5 10   4 1 (0.4000000 0.6000000) *
##                         4115) can>=0.5 19   8 1 (0.4210526 0.5789474) *
##                     1029) without>=0.5 12   4 1 (0.3333333 0.6666667) *
##                    515) make>=0.5 32  13 1 (0.4062500 0.5937500) *
##                129) size>=0.5 11   3 1 (0.2727273 0.7272727) *
##               65) work>=0.5 14   4 1 (0.2857143 0.7142857) *
##             33) built>=0.5 13   4 1 (0.3076923 0.6923077) *
##           17) simpl>=0.5 10   2 1 (0.2000000 0.8000000) *
##          9) style>=0.5 10   2 1 (0.2000000 0.8000000) *
##        5) materi>=0.5 10   2 1 (0.2000000 0.8000000) *
##      3) solut>=0.5 11   2 1 (0.1818182 0.8181818) *
```

```
printcp(DTmodel)
```

```
##
## Classification tree:
## rpart(formula = deal ~ ., data = train_new, method = "class",
##     control = r.ctrl)
##
## Variables actually used in tree construction:
##  [1] built   can     custom  make    materi  product simpl   size
##  [9] solut   style   use     without work
##
## Root node error: 173/346 = 0.5
##
## n= 346
##
##          CP nsplit rel error  xerror     xstd
## 1 0.0404624      0   1.00000 1.16763 0.053000
## 2 0.0346821      1   0.95954 1.20231 0.052649
## 3 0.0317919      4   0.85549 1.15607 0.053102
## 4 0.0289017      6   0.79191 1.10983 0.053435
## 5 0.0173410      7   0.76301 0.95954 0.053716
## 6 0.0086705     10   0.70520 0.93642 0.053652
## 7 0.0000000     13   0.67630 0.95954 0.053716
```

```
plotcp(DTmodel)
```




```
attributes(DTmodel)
```

```
## $names
##  [1] "frame"              "where"              "call"
##  [4] "terms"              "cptable"            "method"
##  [7] "parms"              "control"            "functions"
## [10] "numresp"            "splits"             "variable.importance"
## [13] "y"                  "ordered"
##
## $xlevels
## named list()
##
## $ylevels
## [1] "0" "1"
##
## $class
## [1] "rpart"
```

```
DTmodel$cptable
```

```
##           CP nsplit rel error    xerror       xstd
## 1 0.04046243      0 1.0000000 1.1676301 0.05299962
## 2 0.03468208      1 0.9595376 1.2023121 0.05264863
## 3 0.03179191      4 0.8554913 1.1560694 0.05310156
## 4 0.02890173      6 0.7919075 1.1098266 0.05343512
## 5 0.01734104      7 0.7630058 0.9595376 0.05371631
## 6 0.00867052     10 0.7052023 0.9364162 0.05365155
## 7 0.00000000     13 0.6763006 0.9595376 0.05371631
```

# Pruning the Tree:

```
ptree = prune(DTmodel, 0.054, "CP")
print(ptree)
```

```
## n= 346
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
## 1) root 346 173 0 (0.5000000 0.5000000) *
```

```
rpart.plot(ptree)
```


```
ptree
```

```
## n= 346
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
## 1) root 346 173 0 (0.5000000 0.5000000) *
```

```
DTmodel$variable.importance
```

```
##    custom     solut     simpl     style      work    materi      size
## 2.3226980 2.3004071 2.2515743 2.1130159 2.0755536 1.9868657 1.9533685
##      make   without     built       can       use   product      inch
## 1.7605756 1.5909127 1.4780442 1.2635346 0.8893483 0.7170680 0.3551579
##   prevent     secur       get     young    easier     found      long
## 0.2965077 0.2668045 0.2539970 0.2251574 0.2200719 0.2113016 0.1986866
##    weight    durabl  made_usa      turn       usa     blend     choos
## 0.1986866 0.1778697 0.1778697 0.1778697 0.1778697 0.1659070 0.1659070
##    direct    select     whole      hous      fire     rescu      tool
## 0.1659070 0.1659070 0.1659070 0.1482538 0.1325761 0.1325761 0.1325761
##   can_buy   preserv     recip      side    design    inflat      anim
## 0.1214209 0.1214209 0.1214209 0.1214209 0.1136957 0.1136957 0.1100360
##       mix     bring       low     natur   respons      wast
## 0.1100360 0.0674480 0.0674480 0.0674480 0.0674480 0.0674480
```

Interpretation: . The Pruned Tree is using only one Variable, easier.
. This may not be the best fit. . The Unpruned Tree uses various features earlier

# Performance Measures

The following model performance measures will be calculated on entire data set to gauge the goodness of the model: . Rank Ordering . KS . Area Under Curve (AUC) . Gini Coefficient . Classification Error

```
#CART validation on test data
predCART = predict(ptree, newdata = test_new, type = "class")
predCART1 = predict(ptree, newdata = test_new, type = "prob")
predCART1
```

```
##          0   1
##   [1,] 0.5 0.5
##   [2,] 0.5 0.5
##   [3,] 0.5 0.5
##   [4,] 0.5 0.5
##   [5,] 0.5 0.5
##   [6,] 0.5 0.5
##   [7,] 0.5 0.5
##   [8,] 0.5 0.5
##   [9,] 0.5 0.5
##  [10,] 0.5 0.5
##  [11,] 0.5 0.5
##  [12,] 0.5 0.5
##  [13,] 0.5 0.5
##  [14,] 0.5 0.5
##  [15,] 0.5 0.5
##  [16,] 0.5 0.5
##  [17,] 0.5 0.5
##  [18,] 0.5 0.5
##  [19,] 0.5 0.5
##  [20,] 0.5 0.5
##  [21,] 0.5 0.5
##  [22,] 0.5 0.5
##  [23,] 0.5 0.5
##  [24,] 0.5 0.5
##  [25,] 0.5 0.5
##  [26,] 0.5 0.5
##  [27,] 0.5 0.5
##  [28,] 0.5 0.5
##  [29,] 0.5 0.5
##  [30,] 0.5 0.5
##  [31,] 0.5 0.5
##  [32,] 0.5 0.5
##  [33,] 0.5 0.5
##  [34,] 0.5 0.5
##  [35,] 0.5 0.5
##  [36,] 0.5 0.5
##  [37,] 0.5 0.5
##  [38,] 0.5 0.5
##  [39,] 0.5 0.5
##  [40,] 0.5 0.5
##  [41,] 0.5 0.5
##  [42,] 0.5 0.5
##  [43,] 0.5 0.5
##  [44,] 0.5 0.5
##  [45,] 0.5 0.5
##  [46,] 0.5 0.5
##  [47,] 0.5 0.5
```

```
##  [48,] 0.5 0.5
##  [49,] 0.5 0.5
##  [50,] 0.5 0.5
##  [51,] 0.5 0.5
##  [52,] 0.5 0.5
##  [53,] 0.5 0.5
##  [54,] 0.5 0.5
##  [55,] 0.5 0.5
##  [56,] 0.5 0.5
##  [57,] 0.5 0.5
##  [58,] 0.5 0.5
##  [59,] 0.5 0.5
##  [60,] 0.5 0.5
##  [61,] 0.5 0.5
##  [62,] 0.5 0.5
##  [63,] 0.5 0.5
##  [64,] 0.5 0.5
##  [65,] 0.5 0.5
##  [66,] 0.5 0.5
##  [67,] 0.5 0.5
##  [68,] 0.5 0.5
##  [69,] 0.5 0.5
##  [70,] 0.5 0.5
##  [71,] 0.5 0.5
##  [72,] 0.5 0.5
##  [73,] 0.5 0.5
##  [74,] 0.5 0.5
##  [75,] 0.5 0.5
##  [76,] 0.5 0.5
##  [77,] 0.5 0.5
##  [78,] 0.5 0.5
##  [79,] 0.5 0.5
##  [80,] 0.5 0.5
##  [81,] 0.5 0.5
##  [82,] 0.5 0.5
##  [83,] 0.5 0.5
##  [84,] 0.5 0.5
##  [85,] 0.5 0.5
##  [86,] 0.5 0.5
##  [87,] 0.5 0.5
##  [88,] 0.5 0.5
##  [89,] 0.5 0.5
##  [90,] 0.5 0.5
##  [91,] 0.5 0.5
##  [92,] 0.5 0.5
##  [93,] 0.5 0.5
##  [94,] 0.5 0.5
##  [95,] 0.5 0.5
##  [96,] 0.5 0.5
```

```
##  [97,] 0.5 0.5
##  [98,] 0.5 0.5
##  [99,] 0.5 0.5
## [100,] 0.5 0.5
## [101,] 0.5 0.5
## [102,] 0.5 0.5
## [103,] 0.5 0.5
## [104,] 0.5 0.5
## [105,] 0.5 0.5
## [106,] 0.5 0.5
## [107,] 0.5 0.5
## [108,] 0.5 0.5
## [109,] 0.5 0.5
## [110,] 0.5 0.5
## [111,] 0.5 0.5
## [112,] 0.5 0.5
## [113,] 0.5 0.5
## [114,] 0.5 0.5
## [115,] 0.5 0.5
## [116,] 0.5 0.5
## [117,] 0.5 0.5
## [118,] 0.5 0.5
## [119,] 0.5 0.5
## [120,] 0.5 0.5
## [121,] 0.5 0.5
## [122,] 0.5 0.5
## [123,] 0.5 0.5
## [124,] 0.5 0.5
## [125,] 0.5 0.5
## [126,] 0.5 0.5
## [127,] 0.5 0.5
## [128,] 0.5 0.5
## [129,] 0.5 0.5
## [130,] 0.5 0.5
## [131,] 0.5 0.5
## [132,] 0.5 0.5
## [133,] 0.5 0.5
## [134,] 0.5 0.5
## [135,] 0.5 0.5
## [136,] 0.5 0.5
## [137,] 0.5 0.5
## [138,] 0.5 0.5
## [139,] 0.5 0.5
## [140,] 0.5 0.5
## [141,] 0.5 0.5
## [142,] 0.5 0.5
## [143,] 0.5 0.5
## [144,] 0.5 0.5
## [145,] 0.5 0.5
```

```
## [146,] 0.5 0.5
## [147,] 0.5 0.5
## [148,] 0.5 0.5
## [149,] 0.5 0.5
```

```
## deciling code
decile <- function(x){
  deciles <- vector(length=10)
  for (i in seq(0.1,1,.1)){
    deciles[i*10] <- quantile(x, i, na.rm=T)
  }
  return (
    ifelse(x<deciles[1], 1,
           ifelse(x<deciles[2], 2,
                  ifelse(x<deciles[3], 3,
                         ifelse(x<deciles[4], 4,
                                ifelse(x<deciles[5], 5,
                                       ifelse(x<deciles[6], 6,
                                              ifelse(x<deciles[7], 7,
                                                     ifelse(x<deciles[8], 8,
                                                            ifelse(x<deciles[9], 9, 10
                                                            ))))))))))
}

## deciling
test_new$deciles <- decile(predCART1[,2])
test_new$deal<- as.numeric(test_new$deal)
```

# Model Performance Measure - Rank Ordering

```
# Rank
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:reshape2':
##
##     dcast, melt
```

```
## The following object is masked from 'package:dendextend':
##
##     set
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
library(scales)
```

```
##
## Attaching package: 'scales'
```

```
## The following object is masked from 'package:syuzhet':
##
##     rescale
```

```
## The following object is masked from 'package:plotrix':
##
##     rescale
```

```
tmp_TS = data.table(test_new)
c_rank <- tmp_TS[, list(
 cnt = length(deal),
 cnt_resp = sum(deal),
cnt_non_resp = sum(test_new$deal== 0)) ,
 by= deciles][order(- deciles)]
c_rank$rrate <- round (c_rank$cnt_resp / c_rank$cnt,2);
c_rank$cum_resp <- cumsum(c_rank$cnt_resp)
c_rank$cum_non_resp <- cumsum(c_rank$cnt_non_resp)
c_rank$cum_rel_resp <- round(c_rank$cum_resp / sum(c_rank$cnt_resp),2);
c_rank$cum_rel_non_resp <- round(c_rank$cum_non_resp / sum(c_rank$cnt_non_resp),2);
c_rank$ks <- abs(c_rank$cum_rel_resp - c_rank$cum_rel_non_resp);

library(scales)
c_rank$rrate <- percent(c_rank$rrate)
c_rank$cum_rel_resp <- percent(c_rank$cum_rel_resp)
c_rank$cum_rel_non_resp <- percent(c_rank$cum_rel_non_resp)

print(c_rank)
```

```
##    deciles cnt cnt_resp cnt_non_resp rrate cum_resp cum_non_resp
## 1:      10 149       78           71 52.0%       78           71
##    cum_rel_resp cum_rel_non_resp ks
## 1:         100%             100%  0
```

Interpretation: . The baseline Response Rate is 54.%. . The KS is above 0%, indicating it to be a not very poor model but still need to anlyisi some other accucracy

```
#install.packages("ROCR")
library(ROCR)
```

```
## Loading required package: gplots
```

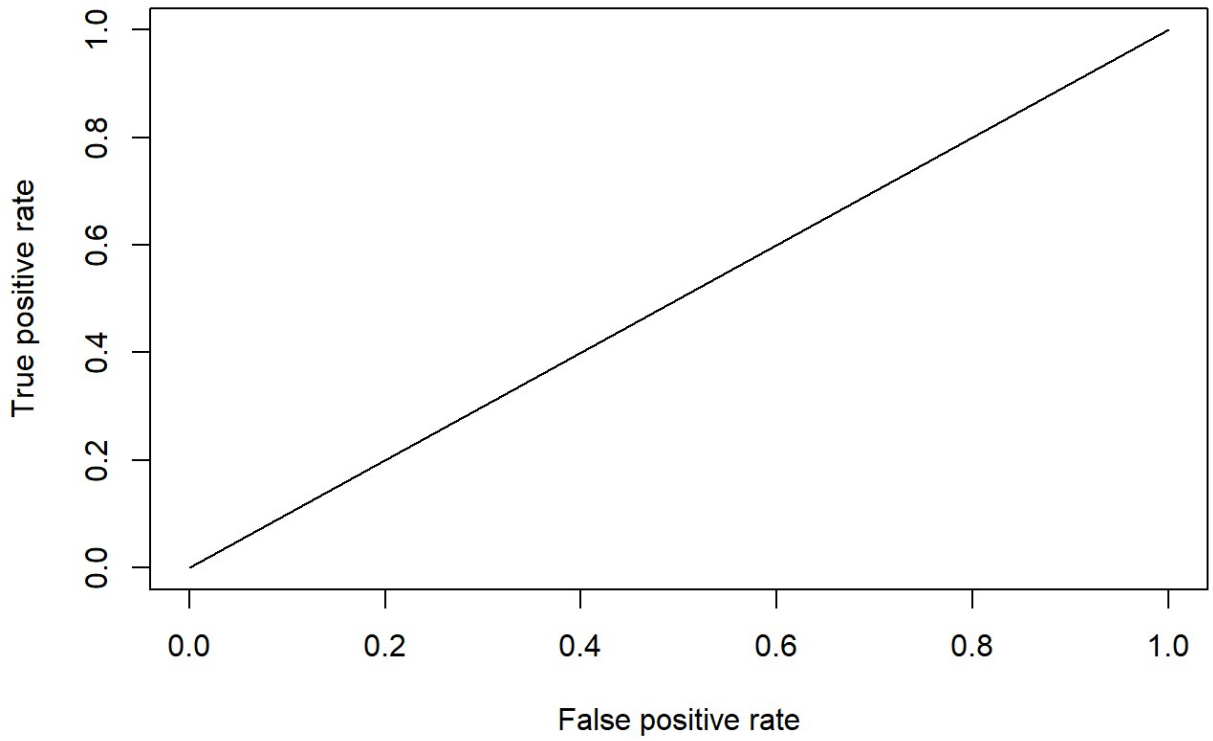
```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:plotrix':
##
##     plotCI
```

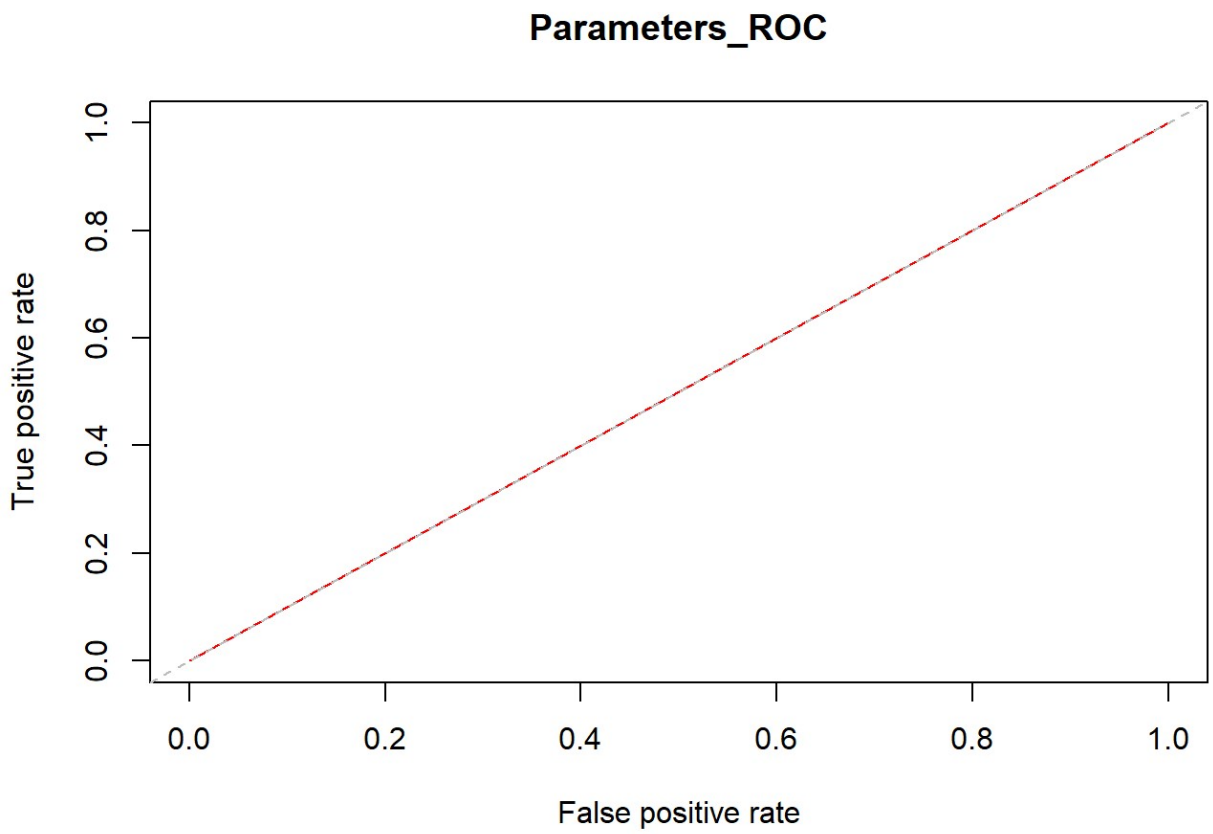
```
## The following object is masked from 'package:wordcloud':
##
##     textplot
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
#Validation on test data
DTpredROC1 = ROCR::prediction(predCART1[,2], test_new$deal)
perf1<-ROCR::performance(DTpredROC1,"tpr","fpr")
plot(perf1)
```




```
plot(perf1,col="red", main="Parameters_ROC")
abline(0,1, lty = 8, col = "grey")
```




Above Graphical representation of the Area Under Curve is as follows:

# Model Performance Measure - KS , Area under Curve & Gini

```
##install.packages("ineq")
library(ineq)
#KS on train
KS <- max(attr(perf1, 'y.values')[[1]]-attr(perf1, 'x.values')[[1]])
KS
```

```
## [1] 0
```

```
auc <- as.numeric(ROCR::performance(DTpredROC1, "auc")@y.values)
auc
```

```
## [1] 0.5
```

```
#gini
gini = ineq(predCART1[,2], type="Gini")
gini
```

```
## [1] 0
```

Above we found The AUC value around 49% indicates the good performance of the model.

## Model Performance Measure - Confusion Matrix

```
with(test_new,table(deal,predCART))
```

```
##     predCART
## deal  0  1
##    0 71  0
##    1 78  0
```

```
(78+2)/(67+2)
```

```
## [1] 1.15942
```

Classification Error Rate = 1- Accuracy = is around 80% The lower the classification error rate, higher the model accuracy, resulting in a better model.
So this not good model accoridng Classification Error Rate

# Random Forest

## The initial build & Optimal No of Trees

```
# Random forest#
dim(train_new)
```

```
## [1] 346 522
```

```
colnames(train_new) <- paste(colnames(train_new), "_c", sep = "")
colnames(test_new) <- paste(colnames(test_new), "_c", sep = "")
colnames(train_new)[1] <- 'deal'
colnames(test_new)[1] <- 'deal'
train_new$deal=factor(train_new$deal)
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```
seed=112
set.seed(seed)
RFmodel = randomForest(deal ~ ., data = train_new, mtry = 7, nodesize = 10, ntree = 50
1, importance = TRUE)
print(RFmodel)
```

```
##
## Call:
##  randomForest(formula = deal ~ ., data = train_new, mtry = 7,      nodesize = 10, n
tree = 501, importance = TRUE)
##                Type of random forest: classification
##                      Number of trees: 501
## No. of variables tried at each split: 7
##
##         OOB estimate of  error rate: 51.73%
## Confusion matrix:
##     0  1 class.error
## 0 105 68   0.3930636
## 1 111 62   0.6416185
```

```
plot(RFmodel, main="")
 legend("topright", c("OOB", "0", "1"), text.col=1:6, lty=1:3, col=1:3)
  title(main="Error Rates Random Forest train_data")
```


```
rf_err_rate <- RFmodel$err.rate
rf_err_rate$ID <- seq.int(nrow(rf_err_rate))
```

```
## Warning in rf_err_rate$ID <- seq.int(nrow(rf_err_rate)): Coercing LHS to a
## list
```

It is observed that as the number of tress increases, the OOB error rate starts decreasing with OOB = 0.076 (the minimum value).

# Variable Importance

To understand the important variables in Random Forest, the following measures are generally used:
Mean Decrease in Accuracy is based on permutation o Randomly permute values of a variable for which importance is to be computed in the OOB sample
Compute the Error Rate with permuted values o Compute decrease in OOB Error rate (Permuted - Not permuted) o Average the decrease over all the trees
Mean Decrease in Gini is computed as “total decrease in node impurities from splitting on the variable, averaged over all trees”

The variables importance is computed as follows:

```
## List the importance of the variables.
impVar <- round(randomForest::importance(RFmodel), 2)
impVar[order(impVar[,1],decreasing = TRUE),]
```

```
##                       0     1 MeanDecreaseAccuracy MeanDecreaseGini
## without_c          8.99  0.19                 6.65             0.82
## size_c             8.23 -1.48                 4.59             0.69
## shape_c            7.67  0.18                 5.28             0.66
## packag_c           7.53  1.38                 5.45             0.72
## start_c            6.90  1.24                 5.03             0.65
## recycl_c           6.63  4.29                 6.38             0.71
## turn_c             6.38  0.30                 4.08             0.75
## can_c              6.31 -5.00                 2.07             0.53
## car_c              6.27  1.40                 4.41             0.53
## roll_c             6.07  2.98                 5.04             0.58
## mani_c             5.95  1.81                 4.49             0.57
## weight_c           5.68  1.91                 4.79             0.48
## work_c             5.44  2.09                 4.78             0.78
## solut_c            5.29  0.45                 3.35             0.66
## clean_c            5.04  0.12                 3.44             0.39
## eas_c              4.93  1.46                 3.71             0.42
## open_c             4.86  2.31                 4.69             0.45
## make_c             4.78 -1.29                 3.34             0.75
## place_c            4.75  0.89                 3.76             0.34
## cook_c             4.69  0.43                 3.13             0.43
## manufactur_c       4.61  1.90                 4.00             0.29
## form_c             4.51  6.61                 6.69             0.66
## regular_c          4.48  0.94                 3.74             0.27
## well_c             4.41  0.53                 3.39             0.33
## whole_c            4.34 -0.94                 1.90             0.35
## name_c             4.34 -1.60                 2.24             0.33
## team_c             4.25  0.02                 2.60             0.29
## goe_c              4.00  1.39                 2.92             0.26
## fire_c             3.92  0.06                 2.27             0.31
## toy_c              3.89  3.15                 4.09             0.45
## great_c            3.86 -3.33                 0.26             0.17
## get_c              3.85 -2.59                 0.51             0.26
## favorit_c          3.78  0.01                 2.08             0.24
## might_c            3.78  0.56                 2.80             0.22
## school_c           3.70  0.69                 2.91             0.29
## built_c            3.68  3.96                 4.91             0.76
## portion_c          3.65 -0.22                 2.24             0.31
## sinc_c             3.64  1.00                 2.35             0.21
## machin_c           3.63  0.54                 2.86             0.23
## just_c             3.63 -4.76                -1.50             0.42
## user_c             3.60  0.17                 2.45             0.55
## bicycl_c           3.48  1.11                 2.76             0.52
## togeth_c           3.34 -0.56                 2.21             0.18
## store_c            3.33 -1.71                 1.34             0.45
## make_fashion_c     3.33  3.00                 3.94             0.26
## promot_c           3.28  0.07                 2.05             0.38
## piec_c             3.26 -0.77                 1.34             0.26
```

```
## retail_c          3.25 -3.34              -0.02            0.37
## inch_c            3.15 -2.12               0.57            0.18
## super_c           3.10 -0.90               1.25            0.29
## tradit_c          3.09  3.46               4.00            0.34
## like_c            3.08 -3.38              -0.52            0.33
## student_c         3.02  2.11               3.04            0.28
## ingredi_c         3.00 -2.51               0.21            0.19
## instead_c         2.93  4.43               4.47            0.66
## choic_c           2.92  0.65               2.01            0.23
## day_c             2.85 -3.12              -0.21            0.33
## young_c           2.84  1.75               2.62            0.20
## support_c         2.84  3.35               3.89            0.22
## origin_c          2.83 -0.97               1.88            0.14
## meal_c            2.81 -1.08               1.02            0.23
## side_c            2.75 -2.04               0.67            0.07
## mission_c         2.75 -2.49               0.00            0.21
## train_c           2.73 -0.36               1.42            0.47
## pay_c             2.73  1.73               2.83            0.15
## see_c             2.66 -1.39               1.06            0.09
## butter_c          2.66  2.17               2.88            0.11
## conveni_c         2.66  0.50               1.99            0.16
## afford_c          2.63 -2.86               0.23            0.22
## rescu_c           2.61 -1.58               0.46            0.15
## ice_cream_c       2.61  1.99               2.76            0.18
## grow_c            2.61  0.08               1.89            0.21
## feel_c            2.57 -2.30              -0.24            0.18
## altern_c          2.54  1.09               2.22            0.46
## shark_tank_c      2.51 -0.36               1.74            0.09
## smart_c           2.50 -1.73               0.70            0.18
## coffe_c           2.48 -0.21               1.40            0.36
## feet_c            2.48  0.15               1.75            0.19
## need_c            2.46 -3.29              -0.47            0.17
## item_c            2.45 -0.73               1.52            0.08
## test_c            2.39 -1.83               0.29            0.16
## versatil_c        2.39 -2.22               0.32            0.26
## typic_c           2.37 -3.26              -0.66            0.21
## educ_c            2.34  5.17               4.76            0.61
## even_c            2.34 -3.59              -1.30            0.29
## card_c            2.32  0.94               2.15            0.36
## caffein_c         2.31 -0.83               1.16            0.16
## effici_c          2.24 -2.04              -0.20            0.20
## shark_c           2.23  1.69               2.33            0.10
## water_c           2.22 -4.34              -1.56            0.42
## chemic_c          2.20 -2.98              -0.80            0.15
## materi_c          2.19  0.29               1.64            0.35
## own_c             2.16  0.70               1.78            0.20
## women_c           2.15  0.21               1.69            0.33
## eco_friend_c      2.11 -1.68               0.47            0.19
## quick_c           2.02 -1.56               0.42            0.15
```

```
## qualiti_c          2.02 -4.56              -1.41            0.30
## tie_c              2.01 -1.95              -0.31            0.14
## portabl_c          1.95 -2.68              -1.18            0.19
## hand_c             1.95 -2.58              -0.37            0.25
## buy_c              1.90 -1.36               0.55            0.24
## ice_c              1.88  1.00               1.97            0.23
## whether_c          1.87 -1.09               0.49            0.11
## base_c             1.85 -1.16               0.46            0.29
## easili_c           1.83 -1.63               0.20            0.27
## connect_c          1.83  0.96               1.77            0.29
## pack_c             1.75  2.95               2.64            0.50
## much_c             1.75 -0.89               0.71            0.07
## creat_c            1.75 -2.10              -0.47            0.26
## durabl_c           1.75  0.25               1.40            0.11
## secret_c           1.75  0.58               1.25            0.15
## household_c        1.74  1.41               1.90            0.07
## serious_c          1.74 -1.00               0.97            0.04
## paper_c            1.74 -1.33              -0.02            0.11
## behind_c           1.74  0.48               1.10            0.08
## pad_c              1.73 -2.44              -0.39            0.18
## fast_c             1.73 -1.41               0.46            0.12
## pocket_c           1.66 -0.51               0.84            0.40
## wed_c              1.64 -0.77               0.78            0.22
## corpor_c           1.64 -0.83               0.58            0.07
## develop_c          1.59 -2.85              -0.73            0.19
## chang_c            1.59 -2.68              -0.83            0.23
## real_c             1.55 -2.63              -0.59            0.15
## season_c           1.52 -0.52               0.65            0.19
## busi_c             1.52 -3.78              -1.71            0.34
## age_c              1.52 -2.99              -1.68            0.10
## long_c             1.51 -1.91              -0.23            0.09
## peopl_c            1.51 -2.79              -0.96            0.27
## blend_c            1.50 -0.92               0.26            0.15
## kid_can_c          1.49 -1.26              -0.26            0.08
## old_c              1.48 -0.32               0.55            0.10
## avail_c            1.47 -4.78              -2.41            0.19
## found_c            1.46 -1.24               0.18            0.13
## websit_c           1.46 -1.14               0.03            0.20
## app_c              1.43 -0.81               0.41            0.35
## wash_c             1.41  1.41               1.63            0.09
## strong_c           1.41  0.00               1.42            0.02
## easi_c             1.37 -2.41              -0.79            0.21
## replac_c           1.37  0.16               1.38            0.07
## recip_c            1.37 -1.99              -0.33            0.16
## right_c            1.36 -1.50               0.03            0.11
## energi_c           1.34 -1.13               0.28            0.25
## save_c             1.34 -0.45               0.50            0.11
## good_c             1.33 -3.74              -2.36            0.17
## featur_c           1.32 -0.07               0.92            0.26
```

```
## readi_c             1.32 -3.26              -2.02            0.07
## can_buy_c           1.32 -2.07              -0.35            0.16
## sound_c             1.32 -2.20              -0.32            0.14
## expand_c            1.29 -2.28              -0.76            0.13
## market_c            1.27 -2.31              -0.80            0.19
## box_c               1.24 -2.04              -0.45            0.17
## natur_c             1.22 -1.80              -0.59            0.24
## natur_ingredi_c     1.18 -2.24              -0.68            0.07
## stop_c              1.16 -0.56               0.76            0.10
## anyon_c             1.15 -1.02               0.01            0.17
## adjust_c            1.12 -1.65              -0.33            0.15
## sturdi_c            1.11 -0.52               0.29            0.06
## help_c              1.10 -0.88               0.09            0.52
## direct_c            1.10  1.74               1.65            0.16
## combin_c            1.06 -1.05               0.10            0.15
## fun_c               1.03 -2.51              -0.96            0.24
## accessori_c         1.03  2.16               2.01            0.45
## simpl_c             1.03  0.62               0.95            0.32
## ever_c              1.02  1.00               1.14            0.08
## cost_c              1.01 -0.37               0.48            0.08
## sugar_c             1.00 -1.73              -0.02            0.08
## award_c             1.00 -0.95              -0.49            0.08
## shirt_c             1.00  1.00               1.42            0.08
## current_c           1.00  0.06               0.42            0.09
## peanut_c            1.00  0.00               1.00            0.03
## duo_c               1.00 -0.38              -0.02            0.09
## sourc_c             0.99  0.50               0.85            0.25
## premium_c           0.98 -0.17               0.65            0.31
## print_c             0.95  1.13               1.34            0.29
## allow_c             0.91  0.99               1.15            0.34
## remov_c             0.90  1.58               1.53            0.43
## five_c              0.88 -0.22               0.22            0.10
## thing_c             0.88 -1.06              -0.26            0.15
## although_c          0.88 -0.06               0.62            0.05
## patent_c            0.87 -0.56               0.12            0.24
## cut_c               0.87 -2.45              -1.00            0.18
## invent_c            0.86  0.71               1.18            0.06
## enjoy_c             0.85 -3.19              -1.92            0.18
## becom_c             0.84 -2.10              -0.83            0.15
## wheel_c             0.84 -1.48              -0.95            0.17
## made_usa_c          0.83 -0.58               0.19            0.10
## focus_c             0.82 -1.41              -0.29            0.17
## sport_c             0.80 -1.12              -0.27            0.19
## skin_c              0.80 -1.83              -0.70            0.41
## special_design_c    0.79 -2.26              -0.99            0.17
## take_c              0.79  2.32               2.10            0.49
## tast_c              0.78 -1.38              -0.21            0.14
## order_c             0.77 -0.97              -0.23            0.26
## consum_c            0.77 -2.72              -1.47            0.10
```

```
## want_c          0.75 -0.35  0.17  0.21
## two_c           0.74 -2.47 -1.10  0.19
## color_c         0.70 -1.47 -0.50  0.42
## vehicl_c        0.69 -0.65  0.14  0.26
## cream_c         0.69  0.99  0.97  0.20
## window_c        0.69 -1.21 -0.30  0.14
## stand_c         0.66 -2.84 -1.51  0.26
## throw_c         0.65 -0.89 -0.23  0.12
## treat_c         0.63 -0.58  0.03  0.34
## heat_c          0.63 -0.84 -0.06  0.20
## power_c         0.61 -2.43 -1.17  0.26
## unlik_c         0.59 -2.37 -1.05  0.12
## produc_c        0.59 -0.75  0.04  0.24
## meet_c          0.59 -0.51  0.20  0.11
## serv_c          0.58  0.14  0.30  0.15
## inspir_c        0.57  0.36  0.56  0.29
## continu_c       0.57 -0.56 -0.03  0.05
## use_c           0.56 -4.33 -2.71  0.26
## damag_c         0.56 -1.70 -0.83  0.10
## appli_c         0.54 -3.74 -2.12  0.20
## companion_c     0.54 -2.00 -1.15  0.07
## varieti_c       0.52  0.04  0.36  0.04
## bring_c         0.48 -2.01 -0.56  0.07
## style_c         0.48 -2.00 -0.94  0.27
## inflat_c        0.46  0.30  0.44  0.20
## parti_c         0.45 -1.03 -0.59  0.14
## put_c           0.44 -1.42 -0.45  0.07
## cloth_c         0.42  0.89  0.85  0.77
## add_c           0.42 -2.25 -1.20  0.22
## everi_c         0.41 -1.80 -0.98  0.12
## organ_c         0.40 -2.15 -1.62  0.20
## parent_c        0.39 -2.58 -1.70  0.15
## bright_c        0.36 -0.53 -0.21  0.07
## low_c           0.34 -1.05 -0.54  0.13
## game_c          0.34 -2.22 -1.23  0.22
## smartphon_c     0.33 -1.78 -0.91  0.14
## set_c           0.32 -2.25 -1.47  0.07
## small_c         0.32 -1.31 -0.64  0.23
## famili_c        0.31 -1.10 -0.42  0.24
## complet_c       0.31  0.01  0.25  0.12
## rather_c        0.30 -2.42 -1.37  0.09
## around_c        0.27 -2.30 -1.37  0.27
## plastic_c       0.27 -1.96 -1.12  0.25
## less_c          0.26 -0.82 -0.36  0.24
## uniqu_c         0.26 -1.77 -1.26  0.26
## year_c          0.24 -1.62 -1.12  0.14
## men_c           0.24  2.42  1.48  0.42
## differ_c        0.21 -1.59 -1.14  0.10
## easier_c        0.18 -0.94 -0.56  0.19
```

```
## gift_c              0.18  0.84                0.68             0.26
## transform_c         0.16  0.07                0.11             0.27
## provid_c            0.13 -2.11               -1.40             0.20
## comfort_c           0.13 -0.39               -0.11             0.31
## bar_c               0.12  1.81                1.26             0.43
## suit_c              0.12 -2.71               -1.62             0.12
## case_c              0.12  0.00                0.04             0.18
## tablet_c            0.11  0.20                0.16             0.14
## free_c              0.10 -1.30               -0.85             0.49
## home_c              0.09 -0.95               -0.62             0.33
## insid_c             0.09 -1.82               -1.13             0.16
## person_c            0.08  0.68                0.35             0.30
## event_c             0.08  0.88                0.73             0.22
## gluten_c            0.08  0.28                0.22             0.15
## detail_c            0.08  0.00               -0.03             0.01
## pitch_c             0.06 -1.73               -1.15             0.05
## sustain_c           0.06  0.55                0.41             0.16
## win_c               0.05 -1.42               -0.71             0.03
## glass_c             0.03  0.83                0.55             0.20
## ultra_c             0.03 -2.92               -1.87             0.14
## someth_c            0.02 -1.94               -1.13             0.07
## care_c              0.01 -2.22               -1.56             0.10
## contain_c           0.01 -2.28               -1.70             0.12
## move_c              0.00  0.13               -0.03             0.08
## surfac_c            0.00  0.00                0.00             0.04
## maintain_c          0.00  0.00                0.00             0.10
## ultim_c             0.00 -1.34               -0.82             0.12
## stay_c              0.00  0.00                0.00             0.05
## struggl_c           0.00  0.00                0.00             0.01
## purchas_c           0.00  0.00                0.00             0.06
## improv_c            0.00  0.00                0.00             0.03
## can_use_c           0.00  0.00                0.00             0.02
## nut_c               0.00  0.00                0.00             0.02
## california_c        0.00  0.00                0.00             0.09
## balanc_c            0.00  0.00                0.00             0.04
## beach_c             0.00 -0.03                0.01             0.03
## storag_c           -0.01  0.38                0.34             0.23
## may_c              -0.01  0.70                0.54             0.07
## pet_c              -0.02  1.41                1.24             0.16
## sold_c             -0.03 -1.14               -0.59             0.25
## special_c          -0.05 -2.91               -1.89             0.23
## expens_c           -0.05  0.63                0.41             0.04
## ball_c             -0.06 -0.58               -0.43             0.23
## point_c            -0.06 -0.67               -0.73             0.11
## new_c              -0.09 -2.06               -1.26             0.26
## dog_c              -0.09  0.13               -0.02             0.26
## kitchen_c          -0.09 -2.47               -1.88             0.10
## ad_c               -0.11 -2.70               -1.75             0.22
## product_c          -0.12 -2.41               -1.66             0.39
```

```
## prevent_c          -0.13 -1.74              -0.89             0.05
## innov_c            -0.13 -2.00              -1.63             0.08
## access_c           -0.14 -1.19              -1.19             0.18
## kind_c             -0.16 -3.44              -2.26             0.18
## belt_c             -0.17 -1.40              -0.98             0.22
## encourag_c         -0.17 -2.64              -1.73             0.11
## allow_user_c       -0.17  1.00               0.62             0.07
## toward_c           -0.20 -1.88              -1.35             0.14
## choos_c            -0.20 -1.21              -0.95             0.13
## first_c            -0.21 -0.74              -0.69             0.06
## drink_c            -0.22 -0.42              -0.55             0.20
## program_c          -0.23 -0.81              -0.54             0.15
## select_c           -0.23 -2.31              -1.78             0.07
## lightweight_c      -0.24 -1.66              -1.33             0.12
## leg_c              -0.25  0.06              -0.05             0.07
## system_c           -0.27 -1.19              -0.96             0.29
## screen_c           -0.30 -0.68              -0.87             0.14
## layer_c            -0.30 -0.52              -0.49             0.10
## onlin_c            -0.33 -1.36              -1.17             0.33
## thank_c            -0.33 -1.65              -1.36             0.09
## build_c            -0.33 -3.35              -2.13             0.15
## mother_c           -0.36 -0.57              -0.59             0.04
## delici_c           -0.36  0.00              -0.51             0.12
## stuf_c             -0.37  0.33              -0.01             0.05
## anoth_c            -0.40 -1.10              -0.99             0.09
## clip_c             -0.40 -1.87              -1.42             0.12
## tool_c             -0.40 -2.67              -2.12             0.16
## experi_c           -0.41 -2.26              -2.07             0.07
## usa_c              -0.42 -2.29              -1.91             0.12
## money_c            -0.42 -0.41              -0.60             0.14
## theme_c            -0.42 -0.62              -0.67             0.11
## live_c             -0.43 -1.65              -1.43             0.20
## devic_c            -0.45 -1.12              -0.95             0.45
## friend_c           -0.45 -2.70              -2.13             0.17
## tri_c              -0.47  0.00              -0.54             0.02
## activ_c            -0.48 -0.78              -0.83             0.22
## book_c             -0.49 -0.33              -0.50             0.14
## yet_c              -0.49 -1.52              -1.29             0.21
## plan_c             -0.49 -0.38              -0.63             0.17
## cours_c            -0.49 -2.07              -1.65             0.18
## deliv_c            -0.50 -2.88              -2.02             0.13
## preserv_c          -0.51 -1.92              -1.47             0.08
## minut_c            -0.52 -0.35              -0.54             0.09
## bottl_c            -0.52  0.59              -0.11             0.14
## chocol_c           -0.54 -0.14              -0.34             0.10
## cup_c              -0.56  0.12              -0.33             0.16
## gear_c             -0.56  1.00              -0.03             0.08
## space_c            -0.59 -2.53              -1.92             0.12
## oil_c              -0.61  0.98               0.61             0.08
```

```
## number_c           -0.61 -0.05              -0.42            0.03
## danc_c             -0.62  0.31              -0.24            0.24
## handl_c            -0.63 -1.00              -1.04            0.06
## made_order_c       -0.68  1.73               0.83            0.26
## prospect_c         -0.68 -2.01              -1.73            0.13
## carri_c            -0.69 -0.25              -0.82            0.11
## learn_c            -0.71 -0.52              -0.99            0.20
## look_c             -0.72 -0.66              -1.05            0.24
## smell_c            -0.73 -0.80              -1.16            0.13
## dress_c            -0.74 -1.14              -1.12            0.10
## eco_c              -0.74 -0.05              -0.43            0.19
## world_c            -0.75  0.89              -0.06            0.33
## perfect_c          -0.76 -0.72              -0.91            0.11
## flat_c             -0.76 -0.43              -0.85            0.07
## alway_c            -0.76 -1.17              -1.17            0.04
## hold_c             -0.77 -0.79              -0.95            0.14
## among_c            -0.78 -2.24              -2.02            0.11
## line_c             -0.79 -1.74              -1.84            0.31
## premium_qualiti_c  -0.80  0.04              -0.49            0.09
## bed_c              -0.81  2.16               0.46            0.43
## wast_c             -0.81 -0.33              -0.60            0.20
## better_c           -0.84 -2.99              -2.49            0.19
## outdoor_c          -0.87 -1.51              -1.71            0.12
## tank_c             -0.89 -2.37              -2.02            0.11
## play_c             -0.92 -2.48              -2.10            0.21
## spray_c            -0.92  1.32              -0.19            0.16
## self_c             -0.94  0.11              -0.58            0.09
## cool_c             -0.94 -1.17              -1.52            0.07
## mix_c              -0.95 -2.66              -2.18            0.30
## larg_c             -0.95 -0.09              -0.60            0.15
## look_like_c        -0.96 -2.23              -1.88            0.07
## secur_c            -0.96 -2.09              -2.30            0.09
## phone_c            -0.96 -1.95              -1.97            0.26
## fit_c              -0.97 -2.29              -2.25            0.24
## brand_c            -0.99 -1.28              -1.48            0.25
## us_c               -1.00 -1.73              -1.66            0.04
## beer_c             -1.00  1.41               0.61            0.09
## pair_c             -1.00 -1.25              -1.39            0.23
## signatur_c         -1.00 -1.42              -1.74            0.01
## simpli_c           -1.00 -2.01              -1.91            0.03
## let_c              -1.00 -0.72              -0.84            0.03
## equip_c            -1.00  1.00               0.03            0.07
## creativ_c          -1.07 -1.27              -1.47            0.15
## non_c              -1.09 -3.39              -2.62            0.21
## reduc_c            -1.10 -1.32              -1.56            0.14
## magic_c            -1.10 -1.47              -1.56            0.26
## call_c             -1.10 -0.39              -0.98            0.23
## partner_c          -1.12 -1.64              -1.62            0.19
## control_c          -1.12 -1.33              -1.75            0.10
```

```
## design_c          -1.13 -2.43              -2.34             0.51
## still_c           -1.13  0.07              -0.57             0.13
## includ_c          -1.14 -1.64              -1.94             0.22
## littl_c           -1.15 -0.02              -0.63             0.07
## protect_c         -1.15 -0.22              -0.99             0.39
## fashion_c         -1.16 -1.76              -2.09             0.27
## effect_c          -1.17 -2.64              -2.40             0.14
## bag_c             -1.18  0.21              -0.72             0.16
## month_c           -1.19 -0.51              -1.02             0.14
## snack_c           -1.20 -1.77              -1.82             0.10
## individu_c        -1.21 -1.80              -1.89             0.18
## flagship_c        -1.21 -0.46              -1.30             0.05
## singl_c           -1.21 -1.11              -1.50             0.21
## bike_c            -1.22 -2.34              -2.02             0.06
## food_c            -1.24 -1.22              -1.62             0.29
## cover_c           -1.24 -1.82              -2.04             0.23
## offer_c           -1.25 -0.58              -1.43             0.31
## receiv_c          -1.26 -1.24              -1.68             0.07
## seat_c            -1.26 -2.76              -2.53             0.20
## across_c          -1.28 -1.78              -1.86             0.09
## big_c             -1.28 -1.64              -1.83             0.08
## classic_c         -1.30 -2.45              -2.07             0.05
## sell_c            -1.30 -1.99              -1.95             0.18
## fresh_c           -1.30  1.68               0.17             0.18
## accord_c          -1.33 -1.76              -1.91             0.08
## video_c           -1.35 -0.63              -1.28             0.20
## technolog_c       -1.35 -1.97              -1.85             0.09
## flavor_c          -1.37 -1.30              -1.77             0.21
## one_c             -1.38 -0.67              -1.54             0.36
## visibl_c          -1.39 -0.79              -1.28             0.08
## process_c         -1.39 -1.89              -1.98             0.14
## practic_c         -1.41 -0.90              -1.28             0.05
## ride_c            -1.41 -1.74              -1.66             0.05
## price_c           -1.41 -3.42              -3.02             0.23
## iphon_c           -1.41 -0.10              -0.82             0.06
## engin_c           -1.42 -0.07              -0.82             0.04
## reusabl_c         -1.46 -1.15              -1.60             0.06
## high_qualiti_c    -1.48 -0.43              -1.54             0.08
## shoe_c            -1.49 -0.89              -1.51             0.19
## lip_c             -1.50 -1.45              -2.10             0.10
## children_c        -1.51  1.07              -0.34             0.54
## content_c         -1.51 -1.37              -1.84             0.14
## electr_c          -1.52 -1.25              -1.73             0.29
## come_c            -1.53 -2.16              -2.34             0.17
## rang_c            -1.54  0.09              -1.20             0.14
## back_c            -1.57 -1.15              -1.79             0.33
## cake_c            -1.57 -0.58              -1.50             0.14
## shop_c            -1.59 -0.48              -1.27             0.19
## display_c         -1.59 -2.25              -2.30             0.10
```

```
## run_c               -1.61 -1.26               -1.90             0.14
## addit_c             -1.61  0.22               -0.98             0.11
## remot_c             -1.61  0.41               -0.89             0.22
## made_c              -1.62 -1.62               -2.39             0.33
## safe_c              -1.63 -1.96               -2.27             0.26
## stick_c             -1.63 -3.10               -2.99             0.21
## top_c               -1.63  0.09               -1.12             0.16
## golf_c              -1.64 -0.62               -1.51             0.09
## air_c               -1.65 -1.29               -1.85             0.09
## wine_c              -1.67 -1.60               -1.97             0.19
## nation_c            -1.68 -0.95               -1.69             0.16
## model_c             -1.70 -1.19               -1.67             0.16
## respons_c           -1.72 -2.76               -2.73             0.10
## audio_c             -1.73  1.00               -1.06             0.07
## best_c              -1.74  1.09               -0.54             0.04
## high_c              -1.75 -1.64               -2.21             0.29
## sale_c              -1.75 -0.13               -1.19             0.14
## mom_c               -1.77  0.97               -0.75             0.08
## travel_c            -1.78 -2.01               -2.64             0.23
## full_c              -1.79 -0.87               -1.62             0.28
## instruct_c          -1.79 -2.85               -2.57             0.12
## love_c              -1.79 -3.09               -3.08             0.11
## fulli_c             -1.81 -2.21               -2.66             0.13
## success_c           -1.85 -0.21               -1.53             0.08
## gourmet_c           -1.88 -3.27               -3.19             0.16
## bodi_c              -1.91 -0.59               -1.44             0.32
## truck_c             -1.91  1.53                0.15             0.21
## center_c            -1.94 -0.65               -1.71             0.14
## babi_c              -1.94 -1.96               -2.53             0.28
## posit_c             -1.96 -1.00               -1.95             0.12
## motorcycl_c         -1.99  0.10               -1.38             0.07
## safeti_c            -2.00 -2.54               -2.80             0.17
## spend_c             -2.00 -1.02               -1.73             0.05
## keep_c              -2.03 -2.32               -2.88             0.22
## custom_c            -2.07  0.53               -1.05             0.53
## find_c              -2.08 -3.19               -3.44             0.17
## flagship_product_c -2.11 -1.27               -2.22             0.07
## go_c                -2.12 -2.31               -3.15             0.23
## music_c             -2.12 -1.64               -2.28             0.09
## hair_c              -2.19 -2.73               -2.96             0.18
## workout_c           -2.32 -0.61               -1.97             0.11
## hous_c              -2.33 -1.83               -2.53             0.13
## ship_c              -2.33 -2.39               -2.69             0.09
## way_c               -2.34 -2.90               -3.86             0.20
## show_c              -2.34 -0.87               -1.87             0.14
## colleg_c            -2.35 -2.66               -2.81             0.10
## anim_c              -2.35 -1.30               -2.12             0.15
## face_c              -2.39 -2.95               -3.57             0.14
## attract_c           -2.40  0.01               -1.70             0.21
```

```
## entertain_c        -2.43 -0.61              -2.05            0.18
## stylish_c          -2.45 -1.47              -2.37            0.06
## resist_c           -2.47 -0.86              -2.11            0.30
## pre_c              -2.50 -1.69              -2.36            0.08
## time_c             -2.54 -0.60              -2.24            0.30
## beverag_c          -2.54 -0.07              -1.81            0.16
## rest_c             -2.57 -1.74              -2.56            0.04
## attach_c           -2.62 -2.40              -3.10            0.21
## healthi_c          -2.65 -1.69              -2.70            0.06
## interchang_c       -2.65 -2.01              -2.66            0.13
## now_c              -2.67 -1.51              -2.68            0.11
## give_c             -2.71 -0.89              -2.46            0.18
## wood_c             -2.71 -1.27              -2.56            0.11
## servic_c           -2.74 -1.32              -2.68            0.31
## room_c             -2.74 -1.19              -2.78            0.13
## mobil_c            -2.81 -3.39              -3.68            0.19
## result_c           -2.83 -2.66              -3.12            0.10
## kit_c              -2.84 -1.62              -2.57            0.09
## band_c             -2.90 -1.79              -2.70            0.16
## light_c            -2.91 -1.71              -2.92            0.38
## countri_c          -2.92 -2.57              -3.09            0.22
## compani_c          -2.95 -2.11              -3.48            0.43
## three_c            -2.98 -0.59              -2.28            0.19
## function_c         -3.01 -0.85              -2.75            0.13
## wearer_c           -3.03 -2.20              -3.34            0.19
## craft_c            -3.04 -2.05              -3.05            0.12
## apparel_c          -3.15  0.15              -2.34            0.27
## profession_c       -3.34 -3.49              -3.83            0.19
## social_c           -3.54 -1.33              -2.97            0.16
## chain_c            -3.79 -2.99              -4.34            0.23
## subscript_c        -3.87 -1.19              -3.36            0.20
## kid_c              -3.93 -2.45              -3.91            0.23
## magnet_c           -4.24 -1.45              -3.42            0.25
```

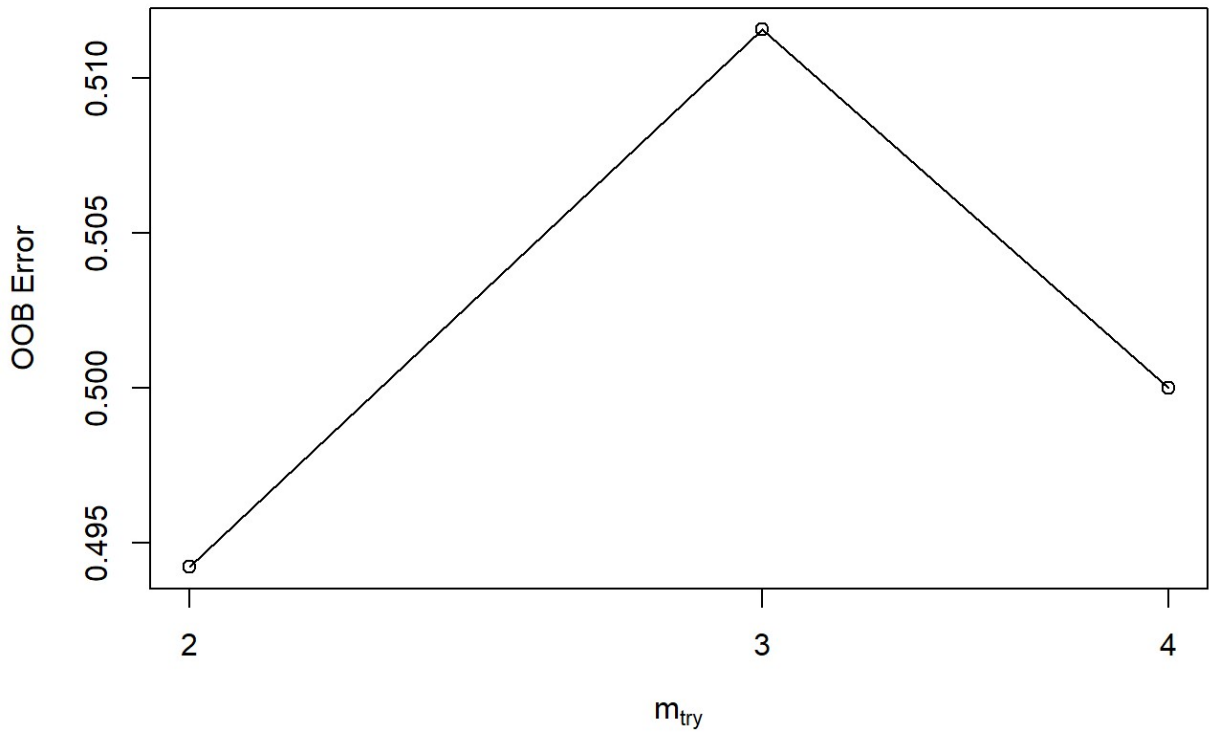
# Optimal value - mtry value

In the random forests literature, the number of variables available for splitting at each tree node is referred to as the mtry parameter. The optimum number of variables is obtained using tuneRF function in test data set as follows

Above mtry = 4 OOB error = 46% (Output: OOB Error Vs Mtry)

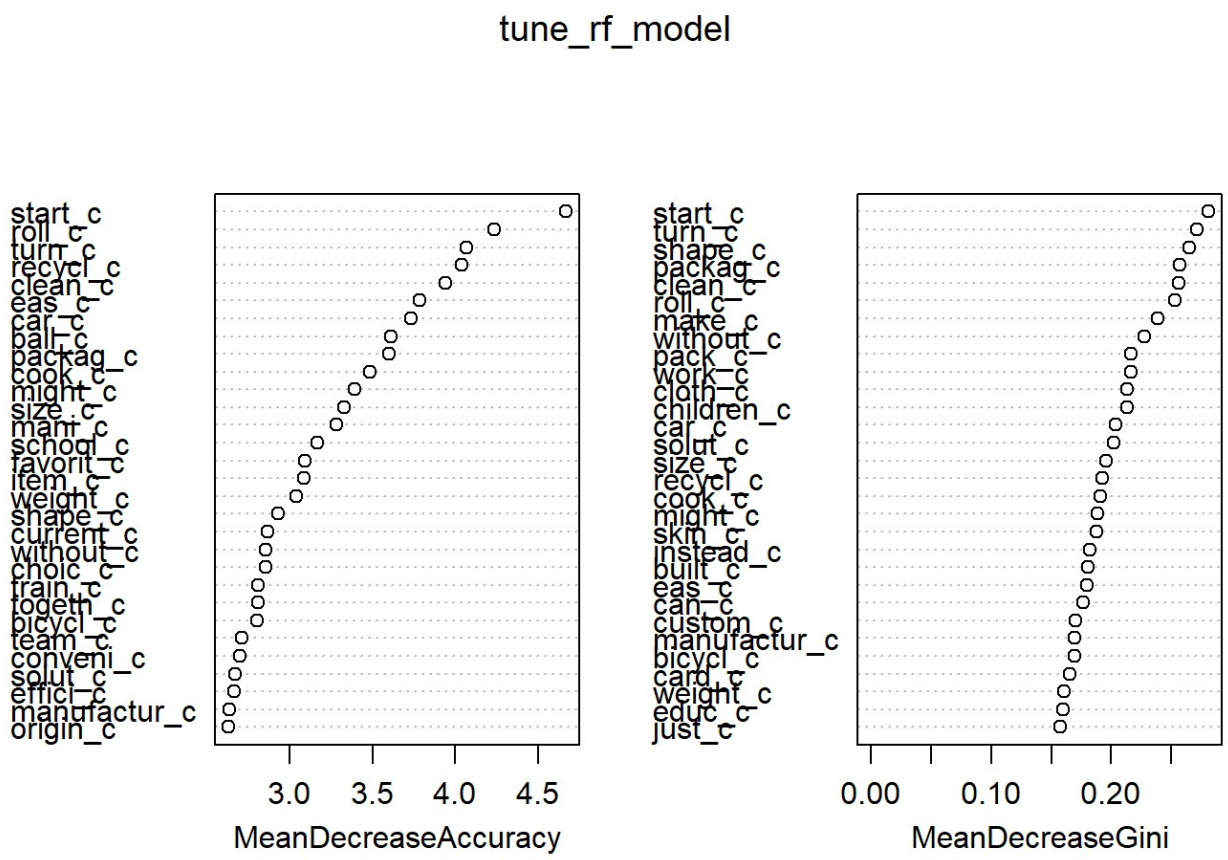
```
tune_rf_model <- tuneRF(x =train_new[,-c(1)],
               y=as.factor(train_new$deal),
               mtryStart = 3,
               ntreeTry= 100,
               stepFactor = 1.5,
               improve = 0.0001,
               trace=TRUE,
               plot = TRUE,
               doBest = TRUE,
               nodesize = 10,
               importance=TRUE
)
```

```
## mtry = 3  OOB error = 51.16%
## Searching left ...
## mtry = 2     OOB error = 49.42%
## 0.03389831 1e-04
## Searching right ...
## mtry = 4     OOB error = 50%
## -0.01169591 1e-04
```



```
#Validate RF model on test data
test_new$predict.class <- predict(tune_rf_model, test_new, type="class")
test_new$predict.score <- predict(tune_rf_model, test_new, type="prob")
```

```
#Checking Variable Importance
varImpPlot(tune_rf_model)
```



```
# deciling
decile <- function(x){
  deciles <- vector(length=10)
  for (i in seq(0.1,1,.1)){
    deciles[i*10] <- quantile(x, i, na.rm=T)
  }
  return (
    ifelse(x<deciles[1], 1,
           ifelse(x<deciles[2], 2,
                  ifelse(x<deciles[3], 3,
                         ifelse(x<deciles[4], 4,
                                ifelse(x<deciles[5], 5,
                                       ifelse(x<deciles[6], 6,
                                              ifelse(x<deciles[7], 7,
                                                     ifelse(x<deciles[8], 8,
                                                            ifelse(x<deciles[9], 9, 10
                                                            ))))))))))
}


test_new$deciles <- decile(test_new$predict.score[,2])
class(test_new$predict.score)
```

```
## [1] "matrix" "votes"
```

```
library(tidyverse)
```

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

```
## v tibble  2.1.3     v purrr   0.3.2
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------- tidyverse_conflicts() --
## x ggplot2::annotate()     masks NLP::annotate()
## x tibble::as_data_frame() masks igraph::as_data_frame(), dplyr::as_data_frame()
## x data.table::between()   masks dplyr::between()
## x readr::col_factor()     masks scales::col_factor()
## x randomForest::combine() masks dplyr::combine()
## x purrr::compose()        masks igraph::compose()
## x igraph::crossing()      masks tidyr::crossing()
## x purrr::discard()        masks scales::discard()
## x Matrix::expand()        masks tidyr::expand()
## x dplyr::filter()         masks stats::filter()
## x data.table::first()     masks dplyr::first()
## x igraph::groups()        masks dplyr::groups()
## x dplyr::lag()            masks stats::lag()
## x data.table::last()      masks dplyr::last()
## x purrr::lift()           masks caret::lift()
## x randomForest::margin()  masks ggplot2::margin()
## x MASS::select()          masks dplyr::select()
## x purrr::simplify()       masks igraph::simplify()
## x purrr::transpose()      masks data.table::transpose()
```

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
```

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

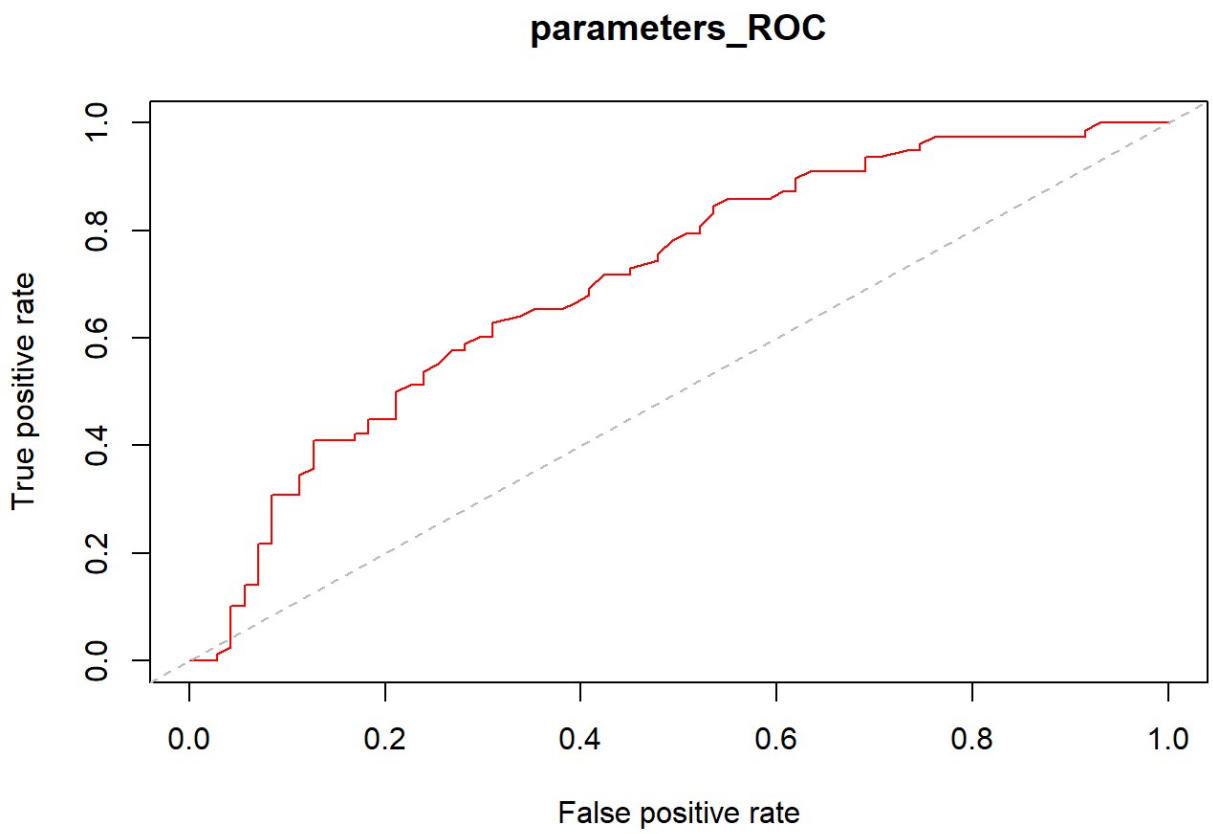
```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
#train_data[, 1] <- as.numeric(as.character( train_data[, 1] ))
#test_data[, 1] <- as.numeric(as.character( test_data[, 1]))
colnames(test_new)[1] <- 'deal'
library(data.table)
library(scales)
```

# Model Performance Measure - Rank Ordering

```
tmp_DT = data.table(test_new)
rank <- tmp_DT[, list(
  cnt = length(deal),
  cnt_resp = sum(deal),
  cnt_non_resp = sum(test_new$deal== 0)) ,
  by= deciles][order(- deciles)]
rank$rrate <- round (rank$cnt_resp / rank$cnt,2);
rank$cum_resp <- cumsum(rank$cnt_resp)
rank$cum_non_resp <- cumsum(rank$cnt_non_resp)
rank$cum_rel_resp <- round(rank$cum_resp / sum(rank$cnt_resp),2);
rank$cum_rel_non_resp <- round(rank$cum_non_resp / sum(rank$cnt_non_resp),2);
rank$ks <- abs(rank$cum_rel_resp - rank$cum_rel_non_resp);
library(scales)
rank$rrate <- percent(rank$rrate)
rank$cum_rel_resp <- percent(rank$cum_rel_resp)
rank$cum_rel_non_resp <- percent(rank$cum_rel_non_resp)

View(rank)
```

```
#test_data[, 1] <- as.numeric(as.character( test_data[, 1]))
#colnames(test_data)[1] <- 'deal'

library(ROCR)
library(ineq)
pred <- prediction(test_new$predict.score[,2], test_new$deal)
perf<-ROCR::performance(pred,"tpr","fpr")
plot(perf,col="red", main="parameters_ROC")
abline(0,1, lty = 8, col = "grey")
```




###Model Performance Measure - KS and Area under Curve

```
KS <- max(attr(perf, 'y.values')[[1]]-attr(perf, 'x.values')[[1]])
KS
```

```
## [1] 0.318346
```

```
## Area Under Curve
auc <-ROCR:: performance(pred,"auc");
auc <- as.numeric(auc@y.values)
auc
```

```
## [1] 0.7100036
```

```
## Gini Coefficient
library(ineq)
gini = ineq(test_new$predict.score[,2], type="Gini")
gini
```

```
## [1] 0.1599411
```

```
## Classification Error
with(test_new,table(deal, predict.class))
```

```
##     predict.class
## deal  0  1
##    0 66  5
##    1 64 14
```

### Logstic Regiressin Model

```
#Logstic regiressin
# partition the data
set.seed(1000)
library(dplyr)
library(caTools)
spl = sample.split(data_shark$deal, SplitRatio=0.70)
train = subset(data_shark, spl ==T)
test = subset(data_shark, spl==F)
dim(train)
```

```
## [1] 347 522
```

# Summary Logitst

```
# create a model
LRmodel = glm(deal ~ ., data = train, family = binomial)
summary(LRmodel)
```

```
##
## Call:
## glm(formula = deal ~ ., family = binomial, data = train)
##
## Deviance Residuals:
##        1         2         5         7         8         9        11
## -0.00001   0.00001  -0.00001  -0.00001  -0.00001  -0.00001   0.00001
##       14        15        16        17        18        21        22
## -0.00001  -0.00001  -0.00001   0.00001  -0.00001   0.00001  -0.00001
##       23        26        27        28        29        30        31
##  0.00001  -0.00001   0.00001  -0.00001   0.00001  -0.00001  -0.00001
##       32        33        34        35        37        39        40
## -0.00001   0.00001   0.00001  -0.00001   0.00001   0.00001  -1.17741
##       41        42        43        45        46        47        50
##  0.00001  -0.00001  -0.00001   0.00001  -0.00001  -0.00001  -0.00001
##       51        52        53        54        55        57        59
##  0.00001  -0.00001   0.00001  -0.00001  -0.00001  -0.00001   0.00001
##       60        61        62        63        64        65        68
##  0.00001   0.00001  -0.00001  -0.00001  -0.00001   0.00001  -0.00001
##       69        70        71        73        74        75        76
##  0.00001  -0.00001   0.00001  -0.00001   0.00001   0.00001  -0.00001
##       77        78        79        80        82        85        88
##  0.00001   0.00001  -0.00001   0.00001   0.00001   0.00001  -0.00001
##       89        90        91        93        94        95        97
##  0.00001  -0.00001  -0.00001  -0.00001  -0.00001   0.00001   0.00001
##       98        99       100       101       102       103       104
## -0.00001   0.00001   0.00001   0.00001   0.00001  -0.00001  -0.00001
##      107       109       110       111       113       115       116
## -0.00001   0.00001  -0.00001  -0.00001  -0.00001   0.00001   0.00001
##      120       121       122       124       125       128       129
##  0.00001  -0.00001   0.00001   0.00001  -0.00001   0.00001   0.00001
##      131       133       137       138       139       141       142
## -0.00001  -0.00001  -0.00001  -0.00001   0.00001  -0.00001  -0.00001
##      143       144       145       146       147       148       150
## -0.00001  -0.00001  -0.00001   0.00001   0.00001  -0.00001  -0.00001
##      151       153       154       155       156       157       158
##  0.00001   0.00001   0.00001  -0.00001   0.00001  -0.00001   0.00001
##      159       160       161       162       163       164       165
## -0.00001   0.00001  -0.00001  -0.00001   0.00001   0.00001  -0.00001
##      166       167       171       173       176       177       178
## -0.00001   0.00001   0.00001  -0.00001  -0.00001   0.00001  -0.00001
##      179       180       181       182       185       186       187
##  0.00001  -0.00001  -0.00001   0.00001  -0.00001   0.00001  -0.00001
##      188       189       190       192       193       194       195
##  0.00001   0.00001  -0.00001  -0.00001  -0.00001   0.00001   0.00001
##      197       198       200       202       203       204       205
##  0.00001  -0.00001  -0.00001  -0.00001   0.00001   0.00001   0.00001
##      206       207       209       211       213       214       215
```

```
## -0.00001  -0.00001   0.00001   0.00001   0.00001  -0.00001  -0.00001
##       216       217       218       219       220       221       223
## -0.00001   0.00001   0.00001  -0.00001   0.00001   0.00001  -0.00001
##       224       225       227       228       229       230       231
## -0.00001  -0.00001  -0.00001  -0.00001  -0.00001   0.00001   0.00001
##       234       236       240       243       244       245       247
##  0.00001   0.00001   0.00001   0.00001   0.00001  -0.00001   0.00001
##       248       250       251       253       254       255       256
## -0.00001   0.00001  -0.00001  -0.00001   0.00001  -0.00001  -0.00001
##       260       261       262       263       264       267       268
## -0.00001   0.00001   0.00001   0.00001  -0.00001   0.00001   0.00001
##       270       271       272       273       274       276       278
## -0.00001  -0.00001  -0.00001  -0.00001   0.00001   0.00001   0.00001
##       279       281       282       283       284       285       286
## -0.00001   0.00001  -0.00001  -0.00001   0.00001   0.00001   0.00001
##       288       289       290       291       292       294       295
## -0.00001   0.00001   0.00001   0.00001  -0.00001  -0.00001   0.00001
##       296       297       298       300       301       302       305
## -0.00001  -0.00001  -0.00001  -0.00001   0.00001   0.00001  -0.00001
##       306       310       312       313       314       317       318
## -0.00001   0.00001  -0.00001  -0.00001   0.00001   0.00001  -0.00001
##       320       322       326       327       328       330       331
##  0.00001   0.00001   0.00001  -0.00001  -0.00001   0.00001  -0.00001
##       333       335       336       337       339       340       341
## -0.00001   0.00001  -0.00001  -0.00001   0.00001   0.00001  -0.00001
##       342       343       344       345       347       348       349
##  0.00001   0.00001   0.00001   0.00001   0.00001   0.00001   0.00001
##       355       356       357       359       360       363       364
##  0.00001   0.00001   0.00001  -0.00001  -0.00001  -0.00001  -0.00001
##       365       366       367       368       369       370       371
## -0.00001   0.00001   0.00001   0.00001  -0.00001  -0.00001   0.00001
##       372       374       376       378       379       380       381
## -0.00001   0.00001   0.00001   0.00001  -0.00001  -0.00001   0.00001
##       382       383       384       385       386       389       390
## -0.00001   0.00001   0.00001  -0.00001   0.00001  -0.00001   0.00001
##       391       393       396       397       398       399       401
## -0.00001   0.00001   0.00001   0.00001  -0.00001  -0.00001  -0.00001
##       402       403       404       405       407       409       410
##  0.00001  -0.00001  -0.00001  -0.00001   0.00001  -0.00001   0.00001
##       415       416       417       418       420       421       422
## -0.00001   0.00001   0.00001   0.00001   0.00001  -0.00001   0.00001
##       423       425       426       427       428       429       430
##  0.00001   0.00001  -0.00001   0.00001  -0.00001   0.00001   0.00001
##       431       432       433       434       435       436       439
## -0.00001  -0.00001   0.00001  -0.00001   0.00001   0.00001  -0.00001
##       443       444       446       447       448       449       450
##  0.00001   1.17741  -0.00001   0.00001   0.00001   0.00001  -0.00001
##       451       452       454       455       456       457       458
## -0.00001  -0.00001   0.00001  -0.00001   0.00001  -0.00001   0.00001
```

```
##      461       463       464       465       466       468       469
##  0.00001  -0.00001   0.00001  -0.00001   0.00001  -0.00001  -0.00001
##      470       471       472       473       474       476       479
## -0.00001   0.00001   0.00001  -0.00001  -0.00001  -0.00001   0.00001
##      480       481       487       488       489       490       491
##  0.00001  -0.00001   0.00001   0.00001  -0.00001   0.00001   0.00001
##      492       493       494       495
##  0.00001   0.00001  -0.00001   0.00001
##
## Coefficients: (176 not defined because of singularities)
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.971e-15  1.414e+00       0        1
## devic            -2.457e+01  1.310e+05       0        1
## retail            1.072e+04  2.832e+07       0        1
## two              -2.427e+04  6.426e+07       0        1
## new               2.854e+03  7.655e+06       0        1
## parent           -2.684e+03  8.188e+06       0        1
## young            -4.963e+04  1.313e+08       0        1
## children         -1.879e+02  5.732e+05       0        1
## make              3.058e+03  8.298e+06       0        1
## easi             -1.037e+03  3.553e+06       0        1
## littl            -1.857e+04  5.186e+07       0        1
## one              -3.477e+03  9.629e+06       0        1
## turn             -1.344e-06  1.853e+05       0        1
## experi           -2.710e+04  7.144e+07       0        1
## play             -2.061e+03  5.439e+06       0        1
## provid           -1.857e+03  4.995e+06       0        1
## posit             1.360e+04  3.655e+07       0        1
## organ            -2.457e+01  3.466e+05       0        1
## pack             -2.215e+04  5.799e+07       0        1
## move             -3.651e+04  9.781e+07       0        1
## servic            3.325e+03  8.763e+06       0        1
## deliv             9.651e+03  2.481e+07       0        1
## colleg            2.102e+04  5.540e+07       0        1
## women             4.562e+03  1.209e+07       0        1
## center           -6.969e+03  1.907e+07       0        1
## room              1.739e+02  2.098e+06       0        1
## offer            -1.791e+03  5.047e+06       0        1
## access           -8.609e+02  2.327e+06       0        1
## educ              9.422e+03  2.662e+07       0        1
## first            -2.457e+01  1.310e+05       0        1
## pitch             6.629e+03  1.774e+07       0        1
## shark            -4.842e+03  1.610e+07       0        1
## tank             -2.457e+01  1.310e+05       0        1
## perfect          -8.568e+03  2.337e+07       0        1
## line              1.554e+03  4.282e+06       0        1
## focus            -4.221e+03  1.154e+07       0        1
## gourmet           3.292e+03  8.368e+06       0        1
## food              4.124e+03  1.090e+07       0        1
```

```
## product          -2.460e+03  6.378e+06       0        1
## sold             -4.071e+03  1.030e+07       0        1
## across           -2.803e+04  7.433e+07       0        1
## store             2.072e+00  1.467e+06       0        1
## includ            1.766e+03  5.032e+06       0        1
## dress             8.377e+03  2.223e+07       0        1
## mani             -2.568e+03  9.121e+06       0        1
## design           -1.399e+03  3.673e+06       0        1
## flavor           -2.457e+01  1.310e+05       0        1
## shark_tank       -6.652e+03  1.972e+07       0        1
## mix               1.423e+03  3.678e+06       0        1
## cloth             5.620e+03  1.519e+07       0        1
## look             -3.082e+03  8.309e+06       0        1
## becom             2.644e+03  6.614e+06       0        1
## big              -1.054e+04  2.797e+07       0        1
## brand            -3.083e+03  8.013e+06       0        1
## activ             2.386e+03  6.325e+06       0        1
## sport             1.021e+03  2.933e+06       0        1
## apparel           2.036e+03  5.387e+06       0        1
## attach            2.435e+03  6.376e+06       0        1
## hold              2.610e+03  6.524e+06       0        1
## side              1.230e+03  3.319e+06       0        1
## screen           -6.300e+03  1.618e+07       0        1
## safeti           -1.120e+04  3.006e+07       0        1
## prevent           7.715e+03  1.992e+07       0        1
## start             3.464e+03  1.029e+07       0        1
## vehicl            2.457e+01  1.310e+05       0        1
## hous             -1.307e+04  3.534e+07       0        1
## get              -3.294e+02  2.213e+06       0        1
## student           1.379e+04  3.593e+07       0        1
## learn             1.843e+04  4.932e+07       0        1
## classic          -3.557e+04  9.364e+07       0        1
## work             -2.144e+03  6.050e+06       0        1
## cook              4.913e+01  1.853e+05       0        1
## household        -6.196e+03  1.659e+07       0        1
## item             -3.469e+03  9.823e+06       0        1
## made              7.240e-07  3.209e+05       0        1
## recycl            2.457e+01  1.310e+05       0        1
## book              5.824e+03  1.584e+07       0        1
## help             -1.476e+03  3.985e+06       0        1
## find             -1.208e+04  3.203e+07       0        1
## fit               1.374e+03  3.677e+06       0        1
## fun               4.381e+03  1.144e+07       0        1
## minut            -1.254e+03  4.957e+06       0        1
## effici           -4.799e+02  1.813e+06       0        1
## entertain        -2.336e+03  6.653e+06       0        1
## us                1.792e+04  4.718e+07       0        1
## set               1.886e+03  5.249e+06       0        1
## card              6.776e+03  1.812e+07       0        1
```

```
## combin          -9.846e+03  2.629e+07       0        1
## keep             3.309e+03  8.897e+06       0        1
## kid              1.249e+03  3.422e+06       0        1
## long            -5.585e+04  1.488e+08       0        1
## school          -1.611e+04  4.228e+07       0        1
## develop          1.123e+04  3.006e+07       0        1
## model            4.379e+03  1.171e+07       0        1
## compani         -3.325e+03  8.751e+06       0        1
## profession      -2.457e+01  1.310e+05       0        1
## remov            7.364e+02  2.123e+06       0        1
## much            -7.825e+03  2.079e+07       0        1
## like            -3.131e+03  8.293e+06       0        1
## yet              2.535e+04  6.687e+07       0        1
## babi            -1.992e+03  5.116e+06       0        1
## spend           -4.245e+03  1.105e+07       0        1
## time             1.903e+03  5.300e+06       0        1
## can              1.640e+03  4.551e+06       0        1
## rescu           -1.089e+04  2.946e+07       0        1
## size            -1.659e+03  4.311e+06       0        1
## quick           -1.930e+03  5.687e+06       0        1
## solut            1.279e+03  3.224e+06       0        1
## add             -1.766e+03  4.942e+06       0        1
## anoth            3.503e+04  9.340e+07       0        1
## layer           -2.081e+03  7.040e+06       0        1
## protect         -5.715e-07  2.620e+05       0        1
## surfac          -7.364e+02  2.172e+06       0        1
## machin           1.685e+03  4.544e+06       0        1
## maintain         1.410e+04  3.790e+07       0        1
## throw            2.339e+04  6.326e+07       0        1
## wash            -6.709e+03  1.869e+07       0        1
## easier          -8.324e+03  2.271e+07       0        1
## tri             -1.979e+04  5.268e+07       0        1
## direct           1.440e+04  3.869e+07       0        1
## bright          -2.112e+04  5.593e+07       0        1
## stylish         -6.378e+03  1.639e+07       0        1
## good            -3.293e+02  1.796e+06       0        1
## own              4.879e+03  1.274e+07       0        1
## coffe            1.267e+03  3.290e+06       0        1
## high             2.981e+03  8.748e+06       0        1
## caffein         -1.442e+04  3.829e+07       0        1
## use             -1.151e+03  3.180e+06       0        1
## toy              9.740e+03  2.557e+07       0        1
## base            -2.743e+03  7.435e+06       0        1
## buy             -3.978e+04  1.044e+08       0        1
## back             1.423e-07  1.853e+05       0        1
## sell             3.009e+03  7.722e+06       0        1
## gift            -4.894e+03  1.228e+07       0        1
## year             1.161e+04  3.158e+07       0        1
## inspir          -1.350e+03  3.681e+06       0        1
```

```
## accessori         7.412e+03  2.014e+07       0        1
## color            -1.162e+04  3.057e+07       0        1
## onlin            -1.485e+03  4.567e+06       0        1
## user              4.913e+01  2.620e+05       0        1
## toward           -9.174e+03  2.462e+07       0        1
## well             -2.571e+03  6.924e+06       0        1
## bar              -1.303e+03  3.228e+06       0        1
## safe              2.894e+03  7.619e+06       0        1
## enjoy             1.504e+03  4.266e+06       0        1
## unlik            -3.153e+04  8.395e+07       0        1
## market           -9.007e+03  2.391e+07       0        1
## low               1.022e+04  2.695e+07       0        1
## less              2.102e+04  5.540e+07       0        1
## sugar            -2.828e+04  7.520e+07       0        1
## damag            -1.905e+04  4.858e+07       0        1
## ingredi           2.273e+03  5.788e+06       0        1
## natur            -1.685e+03  4.610e+06       0        1
## test              1.181e+04  3.125e+07       0        1
## ultim            -4.749e+03  1.269e+07       0        1
## just             -2.314e+01  9.101e+05       0        1
## band             -2.457e+01  1.310e+05       0        1
## weight            2.256e+03  6.700e+06       0        1
## handl            -2.634e+04  6.965e+07       0        1
## famili            2.457e+01  1.310e+05       0        1
## fashion          -3.033e+03  8.280e+06       0        1
## engin            -2.457e+01  1.310e+05       0        1
## bodi             -2.457e+01  1.310e+05       0        1
## order            -3.705e+03  1.114e+07       0        1
## energi            5.645e+03  1.521e+07       0        1
## whole            -5.911e+04  1.570e+08       0        1
## custom           -2.457e+01  3.466e+05       0        1
## choos             1.239e+03  4.564e+06       0        1
## go               -5.620e+03  1.523e+07       0        1
## person           -7.483e+03  2.119e+07       0        1
## made_order       -2.592e+03  8.777e+06       0        1
## natur_ingredi     6.680e+04  1.767e+08       0        1
## beer             -1.254e+03  3.320e+06       0        1
## bottl            -5.399e+03  1.443e+07       0        1
## award             2.764e+04  7.247e+07       0        1
## win              -2.816e+04  7.375e+07       0        1
## materi            3.436e+03  8.831e+06       0        1
## around            1.211e+03  3.125e+06       0        1
## leg              -3.273e+03  8.457e+06       0        1
## program           3.849e+03  1.098e+07       0        1
## danc             -5.247e+03  1.460e+07       0        1
## live              4.936e+03  1.361e+07       0        1
## attract           8.316e+03  2.251e+07       0        1
## five             -8.448e+03  2.246e+07       0        1
## featur           -2.061e+03  5.522e+06       0        1
```

```
## stick            -4.099e+03  1.089e+07       0        1
## free             -2.457e+01  2.929e+05       0        1
## pair              2.524e+04  6.690e+07       0        1
## three            -2.592e+02  1.373e+06       0        1
## instead          -4.250e+02  1.795e+06       0        1
## uniqu             2.061e+03  5.568e+06       0        1
## belt              2.361e+03  6.292e+06       0        1
## allow             1.987e+03  5.387e+06       0        1
## stuf              3.466e+03  1.269e+07       0        1
## anim             -1.648e+04  4.571e+07       0        1
## look_like         3.991e+04  1.061e+08       0        1
## special           4.050e+03  1.088e+07       0        1
## special_design   -7.565e+03  2.038e+07       0        1
## beverag          -3.333e+03  8.888e+06       0        1
## dog              -4.148e+03  1.090e+07       0        1
## golf             -2.529e+03  6.907e+06       0        1
## need              2.900e+03  7.621e+06       0        1
## cover             6.137e-07  2.620e+05       0        1
## run               1.561e+03  4.643e+06       0        1
## replac           -3.903e+03  1.097e+07       0        1
## full              6.432e+02  2.797e+06       0        1
## chain            -9.826e+01  1.186e+06       0        1
## blend             3.097e+03  8.357e+06       0        1
## season           -1.024e+04  2.750e+07       0        1
## portabl          -2.188e+04  5.738e+07       0        1
## ball              2.438e+04  6.424e+07       0        1
## patent            9.132e+03  2.452e+07       0        1
## place            -2.818e+03  8.023e+06       0        1
## motorcycl        -6.677e+02  2.800e+06       0        1
## kind             -2.829e+03  7.654e+06       0        1
## eco              -7.259e+02  6.571e+06       0        1
## friend           -1.936e+02  6.435e+05       0        1
## eco_friend        1.942e+03  7.881e+06       0        1
## individu         -5.179e+03  1.443e+07       0        1
## packag           -1.443e+03  3.772e+06       0        1
## clean             2.546e+03  6.842e+06       0        1
## easili            4.000e+03  1.079e+07       0        1
## bicycl            1.508e+03  4.717e+06       0        1
## face              1.423e+03  3.650e+06       0        1
## storag            1.037e+03  2.798e+06       0        1
## pocket           -2.484e+03  6.404e+06       0        1
## shop              1.228e+02  1.179e+06       0        1
## way              -1.211e+03  3.152e+06       0        1
## save              3.395e+04  9.048e+07       0        1
## mobil            -2.272e-07  1.853e+05       0        1
## ice               2.500e+04  6.601e+07       0        1
## cream            -1.970e+04  5.275e+07       0        1
## busi             -1.254e+03  3.235e+06       0        1
## ice_cream        -4.074e+03  1.298e+07       0        1
```

```
## ever            -1.051e+04  2.894e+07       0        1
## whether          4.579e+03  1.152e+07       0        1
## regular         -3.688e+04  9.728e+07       0        1
## stand           -3.608e+04  9.655e+07       0        1
## print            2.501e+02  2.333e+06       0        1
## see              1.966e+04  5.398e+07       0        1
## sinc             1.369e+04  3.665e+07       0        1
## take            -2.457e+01  1.310e+05       0        1
## still            3.455e+03  1.060e+07       0        1
## even             2.147e+04  5.685e+07       0        1
## show             5.015e+03  1.337e+07       0        1
## manufactur      -2.121e+04  5.664e+07       0        1
## cost             2.487e+04  6.836e+07       0        1
## age             -1.225e+03  3.416e+06       0        1
## shape            3.775e+03  1.096e+07       0        1
## creat           -8.945e+03  2.375e+07       0        1
## encourag        -9.843e+03  2.560e+07       0        1
## pre              3.485e+03  9.317e+06       0        1
## plan             6.705e+03  1.831e+07       0        1
## usa             -1.004e+04  2.645e+07       0        1
## made_usa                NA         NA      NA       NA
## hand             8.655e-07  5.558e+05       0        1
## want            -1.062e+04  2.815e+07       0        1
## thing           -2.619e+03  7.346e+06       0        1
## practic          1.238e+04  3.302e+07       0        1
## receiv          -4.415e+04  1.184e+08       0        1
## give            -5.953e+03  1.561e+07       0        1
## home             5.196e-08  1.853e+05       0        1
## shoe            -2.930e+03  8.092e+06       0        1
## sale            -1.031e+04  2.814e+07       0        1
## parti            1.283e+04  3.485e+07       0        1
## stay            -7.734e+03  2.087e+07       0        1
## without         -4.771e+03  1.277e+07       0        1
## struggl          1.504e+04  4.089e+07       0        1
## plastic         -6.776e+03  1.806e+07       0        1
## men              6.171e+03  1.658e+07       0        1
## signatur         4.986e+04  1.328e+08       0        1
## power           -7.538e+02  2.114e+06       0        1
## secur           -4.078e+03  1.052e+07       0        1
## strong          -1.977e+04  5.234e+07       0        1
## magnet           8.029e+02  2.167e+06       0        1
## shirt           -1.399e+03  3.671e+06       0        1
## produc           2.435e+03  6.393e+06       0        1
## expand          -5.701e+03  1.516e+07       0        1
## suit             3.231e+04  8.352e+07       0        1
## money            4.026e+03  1.068e+07       0        1
## clip            -1.833e+03  4.981e+06       0        1
## tie             -1.159e+04  3.098e+07       0        1
## travel          -3.325e+03  8.774e+06       0        1
```

```
## bag              -1.047e+03  2.644e+06       0        1
## premium           2.071e+04  5.558e+07       0        1
## wine              4.913e+01  1.853e+05       0        1
## glass            -2.212e+04  5.848e+07       0        1
## cup               1.737e+04  4.554e+07       0        1
## flagship          4.169e+03  1.088e+07       0        1
## origin            7.147e+03  1.912e+07       0        1
## flagship_product         NA         NA      NA       NA
## subscript        -1.304e+04  3.431e+07       0        1
## fresh            -1.228e+02  2.929e+05       0        1
## bring             1.004e-06  4.143e+05       0        1
## self              9.712e+03  2.533e+07       0        1
## air               5.592e+03  1.521e+07       0        1
## come              2.206e+03  6.018e+06       0        1
## reduc            -1.357e+04  3.655e+07       0        1
## effect            1.732e+04  4.488e+07       0        1
## super            -2.416e+04  6.565e+07       0        1
## lightweight       7.699e+03  2.056e+07       0        1
## ultra             4.657e+04  1.229e+08       0        1
## durabl            3.208e+03  8.703e+06       0        1
## ride              3.646e+02  1.061e+06       0        1
## tradit           -5.641e+03  1.527e+07       0        1
## simpli            3.670e+04  9.886e+07       0        1
## bike              1.779e+04  4.643e+07       0        1
## feet              1.116e+04  3.027e+07       0        1
## fulli             1.301e+04  3.490e+07       0        1
## readi            -2.959e+04  8.008e+07       0        1
## resist           -2.386e+03  6.337e+06       0        1
## mom               1.374e+03  3.671e+06       0        1
## qualiti           2.016e+03  5.119e+06       0        1
## care             -3.823e+03  1.200e+07       0        1
## craft            -2.457e+01  1.310e+05       0        1
## piec              2.824e+04  7.473e+07       0        1
## serv              1.734e+04  4.622e+07       0        1
## found             1.762e+03  5.032e+06       0        1
## current          -5.102e+04  1.343e+08       0        1
## expens            3.026e+04  8.304e+07       0        1
## sourc            -2.244e+04  5.967e+07       0        1
## sustain           2.769e+04  7.310e+07       0        1
## wood              1.861e+04  4.798e+07       0        1
## high_qualiti      1.272e+04  3.322e+07       0        1
## avail             3.349e+03  8.776e+06       0        1
## choic            -9.817e+03  2.671e+07       0        1
## let               1.154e+04  3.095e+07       0        1
## kid_can          -3.561e+04  9.284e+07       0        1
## anyon            -9.915e+03  2.673e+07       0        1
## pet               2.112e+03  5.435e+06       0        1
## smell            -7.457e+03  1.976e+07       0        1
## differ            4.684e+04  1.244e+08       0        1
```

```
## spray             2.027e+02  6.777e+05       0        1
## chemic            9.275e+03  2.638e+07       0        1
## continu          -2.390e+02  1.249e+06       0        1
## can_buy           3.238e+04  8.502e+07       0        1
## countri           2.983e+03  7.438e+06       0        1
## style            -4.874e+03  1.247e+07       0        1
## truck            -1.665e+03  4.552e+06       0        1
## stop              7.916e+03  2.111e+07       0        1
## seat              1.102e+03  2.891e+06       0        1
## addit            -1.641e+04  4.406e+07       0        1
## preserv           4.514e+02  1.295e+06       0        1
## lip              -4.445e+03  1.225e+07       0        1
## light             2.457e+01  1.310e+05       0        1
## car                      NA         NA      NA       NA
## window                   NA         NA      NA       NA
## small                    NA         NA      NA       NA
## case                     NA         NA      NA       NA
## gear                     NA         NA      NA       NA
## skin                     NA         NA      NA       NA
## appli                    NA         NA      NA       NA
## built                    NA         NA      NA       NA
## cake                     NA         NA      NA       NA
## recip                    NA         NA      NA       NA
## ship                     NA         NA      NA       NA
## system           -2.457e+01  1.310e+05       0        1
## pad                      NA         NA      NA       NA
## train                    NA         NA      NA       NA
## video                    NA         NA      NA       NA
## game                     NA         NA      NA       NA
## altern                   NA         NA      NA       NA
## cours                    NA         NA      NA       NA
## among                    NA         NA      NA       NA
## non               2.457e+01  1.310e+05       0        1
## wed                      NA         NA      NA       NA
## event                    NA         NA      NA       NA
## fast                     NA         NA      NA       NA
## fire                     NA         NA      NA       NA
## connect                  NA         NA      NA       NA
## tool                     NA         NA      NA       NA
## kit                      NA         NA      NA       NA
## box                      NA         NA      NA       NA
## success                  NA         NA      NA       NA
## healthi                  NA         NA      NA       NA
## month                    NA         NA      NA       NA
## display                  NA         NA      NA       NA
## interchang               NA         NA      NA       NA
## allow_user               NA         NA      NA       NA
## theme                    NA         NA      NA       NA
## make_fashion             NA         NA      NA       NA
```

```
## water             4.913e+01  1.853e+05       0        1
## contain                  NA         NA      NA       NA
## top                      NA         NA      NA       NA
## eas                      NA         NA      NA       NA
## websit                   NA         NA      NA       NA
## peopl                    NA         NA      NA       NA
## purchas                  NA         NA      NA       NA
## partner                  NA         NA      NA       NA
## larg                     NA         NA      NA       NA
## prospect                 NA         NA      NA       NA
## transform                NA         NA      NA       NA
## carri                    NA         NA      NA       NA
## bed                      NA         NA      NA       NA
## magic                    NA         NA      NA       NA
## day                      NA         NA      NA       NA
## electr                   NA         NA      NA       NA
## right                    NA         NA      NA       NA
## music                    NA         NA      NA       NA
## real                     NA         NA      NA       NA
## favorit                  NA         NA      NA       NA
## instruct                 NA         NA      NA       NA
## open                     NA         NA      NA       NA
## point                    NA         NA      NA       NA
## serious                  NA         NA      NA       NA
## butter                   NA         NA      NA       NA
## oil                      NA         NA      NA       NA
## gluten                   NA         NA      NA       NA
## now                      NA         NA      NA       NA
## name                     NA         NA      NA       NA
## insid                    NA         NA      NA       NA
## inflat                   NA         NA      NA       NA
## chang                    NA         NA      NA       NA
## better                   NA         NA      NA       NA
## comfort                  NA         NA      NA       NA
## adjust                   NA         NA      NA       NA
## varieti                  NA         NA      NA       NA
## best                     NA         NA      NA       NA
## space                    NA         NA      NA       NA
## flat                     NA         NA      NA       NA
## creativ                  NA         NA      NA       NA
## chocol                   NA         NA      NA       NA
## corpor                   NA         NA      NA       NA
## improv                   NA         NA      NA       NA
## equip                    NA         NA      NA       NA
## technolog                NA         NA      NA       NA
## although                 NA         NA      NA       NA
## number                   NA         NA      NA       NA
## might                    NA         NA      NA       NA
## conveni                  NA         NA      NA       NA
```

```
## rest                NA          NA          NA          NA
## alway               NA          NA          NA          NA
## rather              NA          NA          NA          NA
## smart               NA          NA          NA          NA
## kitchen             NA          NA          NA          NA
## treat               NA          NA          NA          NA
## cut                 NA          NA          NA          NA
## paper               NA          NA          NA          NA
## accord              NA          NA          NA          NA
## can_use             NA          NA          NA          NA
## price               NA          NA          NA          NA
## togeth              NA          NA          NA          NA
## everi               NA          NA          NA          NA
## rang                NA          NA          NA          NA
## premium_qualiti     NA          NA          NA          NA
## meal                NA          NA          NA          NA
## portion             NA          NA          NA          NA
## wast                NA          NA          NA          NA
## workout             NA          NA          NA          NA
## secret              NA          NA          NA          NA
## visibl              NA          NA          NA          NA
## wearer              NA          NA          NA          NA
## mother              NA          NA          NA          NA
## drink               NA          NA          NA          NA
## put                 NA          NA          NA          NA
## thank               NA          NA          NA          NA
## wheel               NA          NA          NA          NA
## inch                NA          NA          NA          NA
## nut                 NA          NA          NA          NA
## singl               NA          NA          NA          NA
## peanut              NA          NA          NA          NA
## social              NA          NA          NA          NA
## call                NA          NA          NA          NA
## smartphon           NA          NA          NA          NA
## build               NA          NA          NA          NA
## app                 NA          NA          NA          NA
## reusabl             NA          NA          NA          NA
## sturdi              NA          NA          NA          NA
## phone               NA          NA          NA          NA
## grow                NA          NA          NA          NA
## afford              NA          NA          NA          NA
## content             NA          NA          NA          NA
## consum              NA          NA          NA          NA
## pay                 NA          NA          NA          NA
## may                 NA          NA          NA          NA
## feel                NA          NA          NA          NA
## form                NA          NA          NA          NA
## ad                  NA          NA          NA          NA
## roll                NA          NA          NA          NA
```

```
## california                NA         NA      NA       NA
## balanc                    NA         NA      NA       NA
## control                   NA         NA      NA       NA
## support                   NA         NA      NA       NA
## cool                      NA         NA      NA       NA
## heat                      NA         NA      NA       NA
## `function`                NA         NA      NA       NA
## remot                     NA         NA      NA       NA
## tast                      NA         NA      NA       NA
## hair                      NA         NA      NA       NA
## outdoor                   NA         NA      NA       NA
## innov                     NA         NA      NA       NA
## invent                    NA         NA      NA       NA
## love                      NA         NA      NA       NA
## sound                     NA         NA      NA       NA
## snack                     NA         NA      NA       NA
## someth                    NA         NA      NA       NA
## result                    NA         NA      NA       NA
## process                   NA         NA      NA       NA
## delici                    NA         NA      NA       NA
## versatil                  NA         NA      NA       NA
## goe                       NA         NA      NA       NA
## simpl                     NA         NA      NA       NA
## complet                   NA         NA      NA       NA
## mission                   NA         NA      NA       NA
## detail                    NA         NA      NA       NA
## old                       NA         NA      NA       NA
## great                     NA         NA      NA       NA
## audio                     NA         NA      NA       NA
## promot                    NA         NA      NA       NA
## iphon                     NA         NA      NA       NA
## nation                    NA         NA      NA       NA
## team                      NA         NA      NA       NA
## duo                       NA         NA      NA       NA
## behind                    NA         NA      NA       NA
## tablet                    NA         NA      NA       NA
## meet                      NA         NA      NA       NA
## select                    NA         NA      NA       NA
## world                     NA         NA      NA       NA
## respons                   NA         NA      NA       NA
## beach                     NA         NA      NA       NA
## typic                     NA         NA      NA       NA
## companion                 NA         NA      NA       NA
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 480.9721  on 346  degrees of freedom
## Residual deviance:   2.7726  on   1  degrees of freedom
## AIC: 694.77
```

```
##
## Number of Fisher Scoring iterations: 23
```

## Validate the model

```
#Validate the model
predTest = predict(LRmodel, newdata = test, type="response")
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading
```

```
table(test$deal, predTest>0.3)
```

```
##
##     FALSE TRUE
##   0    44   29
##   1    43   32
```

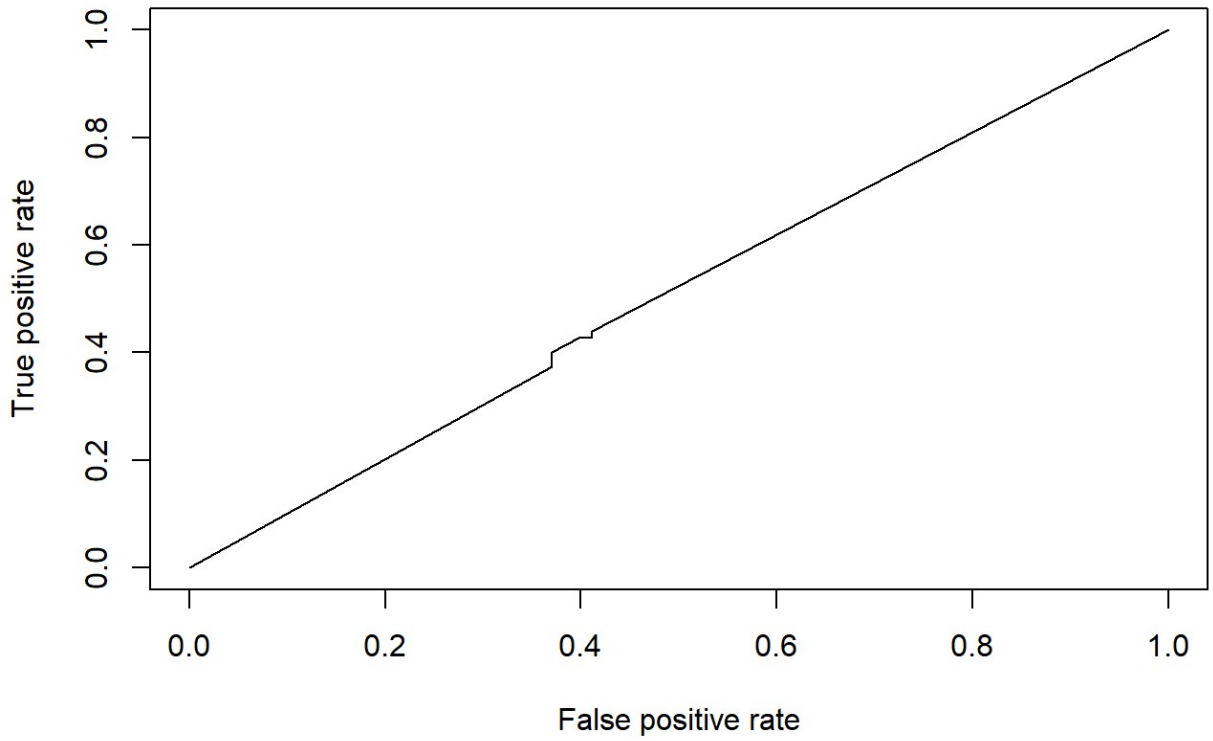
```
(73+75)/nrow(na.omit(test))
```

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

```
library(ROCR)
ROCRpred = prediction(predTest, test$deal)
as.numeric(performance(ROCRpred, "auc")@y.values)
```

```
## [1] 0.5103196
```

```
perf = performance(ROCRpred, "tpr","fpr")
plot(perf)
```




# Adding Ratio in the data set

Now let's add additional variable called as Ratio which will be derived using column askfor/valuation and then we will re-run the models to see if we can have improved accuracy in the models

```
#Adding Ratio in the data set#
names(Shark)
```

```
##  [1] "deal"                  "description"
##  [3] "episode"               "category"
##  [5] "entrepreneurs"         "location"
##  [7] "website"               "askedFor"
##  [9] "exchangeForStake"      "valuation"
## [11] "season"                "shark1"
## [13] "shark2"                "shark3"
## [15] "shark4"                "shark5"
## [17] "title"                 "episode.season"
## [19] "Multiple.Entreprenuers"
```

```
ratio_data <- subset(Shark, select = c(8,10))
ratio <- as.integer(ratio_data$askedFor/ratio_data$valuation*100)
ratio
```

```
##   [1]  14  10  15  25  10  15  20  20  10  10  35  10  20   5  10  15  15
##  [18]  40  30  25  20  25  20  25  29  20  15  25  10  25  70  20  25  25
##  [35]  25  25  10  20  20  10  10  15  17  50  10  25  20  10  10  40  20
##  [52]  20  40  32  29  30  25  15  25  50  25  10  20  15  10   5   5  25
##  [69]  10  20  10  20  15  25  20  10  25  25  20  20  20   5  29  30  25
##  [86]  20  25  20  25  10  15  20  25  14  15  40  14  15  20  45   5  15
## [103]  10  10  25  40  10  20  10   5  50  20  25  35  30  25  20  25  20
## [120]  15  25  25  20  15  50  25  15  20  12  50  15  20  25   5  10  30
## [137]  15  15  15  51  10  20   5   2  20  20  14  30  30  20  32  14  20
## [154]   5  20  10  30  10  10  32  10  10   7  10  32  10  10   5  14  15
## [171]  20  10  12  15  33  10  25  40  10  25  14  14   5  10  20  10  20
## [188]  10   5  14  25  15  10  14  20  25  10  10  20  20  30   5  15  25
## [205]  15  33  29  10  12  12   5  20  26  14   8 100  10  15  20  12  14
## [222]  30  25  35   5  10  20  10  20  20  30  20  10  25  20  15  40  10
## [239]  25  10  14  10  20  20  22  20  20  40  15  20  25  10  30  20  34
## [256]  10   5  25  25  10  14  10  25  20  20  10   5  10  33  20  10  32
## [273]  20  10  15  14   5  25  10   6  15  25  20  10  10  20  20  20  20
## [290]  15   6  20  20  12  20  10  15  15  10  20  15  12  10  10  30  15
## [307]  15  15  20   5  15  10  20  20  10  20  10  30  10  10  20   5  29
## [324]   5  10  20  10   5  25  14  15  10   5  20  10  20  25  10  25  25
## [341]  20  20  17  29  10  20   5  30  10   4  10  29  15  15  20  10  20
## [358]  10  10  10  15  14  15  10  20  20  10  10  10   3  29  25  11  20
## [375]  35  20  10  15  20  20  10  20   5   5  20  15  15  10   5  10  30
## [392]  25   5  25  25   5   5  20  20  10  35  10  10   5  14   5  10  29
## [409]  15  10  20  20  10   5  10  20  33   5  14  10  20  29  20  10  10
## [426]  20  20  20  11  10   5  10  20  15  30  20   8   6  10  20   5  10
## [443]   7  15  10  20  10  20  15  20   5  10  15  20   5  14  10  10  15
## [460]  10  20  10  15   5   5  12  20  20  13   5  10  20   5  10  15  10
## [477]  20   5  25  10  17  10   8  20   3  20  50  10  14  10  20  10   5
## [494]  20  10
```

```
str(data_shark)
```

```
## 'data.frame':    495 obs. of  522 variables:
##  $ deal            : num  0 1 1 0 0 1 0 0 0 1 ...
##  $ devic           : num  1 0 0 0 0 0 0 0 1 0 ...
##  $ retail          : num  0 2 0 0 0 1 0 0 0 0 ...
##  $ two             : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ new             : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ parent          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ young           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ children        : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ make            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ easi            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ littl           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ one             : num  0 0 1 0 0 1 0 0 0 0 ...
##  $ turn            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ experi          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ play            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ provid          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ posit           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ organ           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ pack            : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ move            : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ servic          : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ deliv           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ colleg          : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ women           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ center          : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ room            : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ offer           : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ access          : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ educ            : num  0 0 0 0 1 0 0 0 0 1 ...
##  $ first           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ pitch           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ shark           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ tank            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ perfect         : num  0 0 0 0 0 2 0 0 0 0 ...
##  $ line            : num  0 0 0 0 0 1 1 0 0 0 ...
##  $ focus           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ gourmet         : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ food            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ product         : num  0 0 0 0 0 2 0 0 0 0 ...
##  $ sold            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ across          : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ store           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ includ          : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ dress           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ mani            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ design          : num  0 0 0 0 0 1 0 0 0 1 ...
##  $ flavor          : num  0 0 0 0 0 1 0 0 0 0 ...
```

```
##  $ shark_tank      : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ mix             : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ cloth           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ look            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ becom           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ big             : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ brand           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ activ           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ sport           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ apparel         : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ attach          : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ hold            : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ side            : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ screen          : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ safeti          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ prevent         : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ start           : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ vehicl          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ hous            : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ get             : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ student         : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ learn           : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ classic         : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ work            : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ cook            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ household       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ item            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ made            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ recycl          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ book            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ help            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ find            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ fit             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ fun             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ minut           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ effici          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ entertain       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ us              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ set             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ card            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ combin          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ keep            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ kid             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ long            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ school          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ develop         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ model           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ compani         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ profession      : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ remov           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ much            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ like            : num  0 0 0 0 0 0 0 0 0 0 ...
##   [list output truncated]
```

```
shark_data2 = cbind(data_shark$deal,ratio,data_shark)
data_shark1=shark_data2[,-c(3)]
names(data_shark1)[names(data_shark1) == "data_shark$deal"] <- "deal"
```

# Rebuild Model

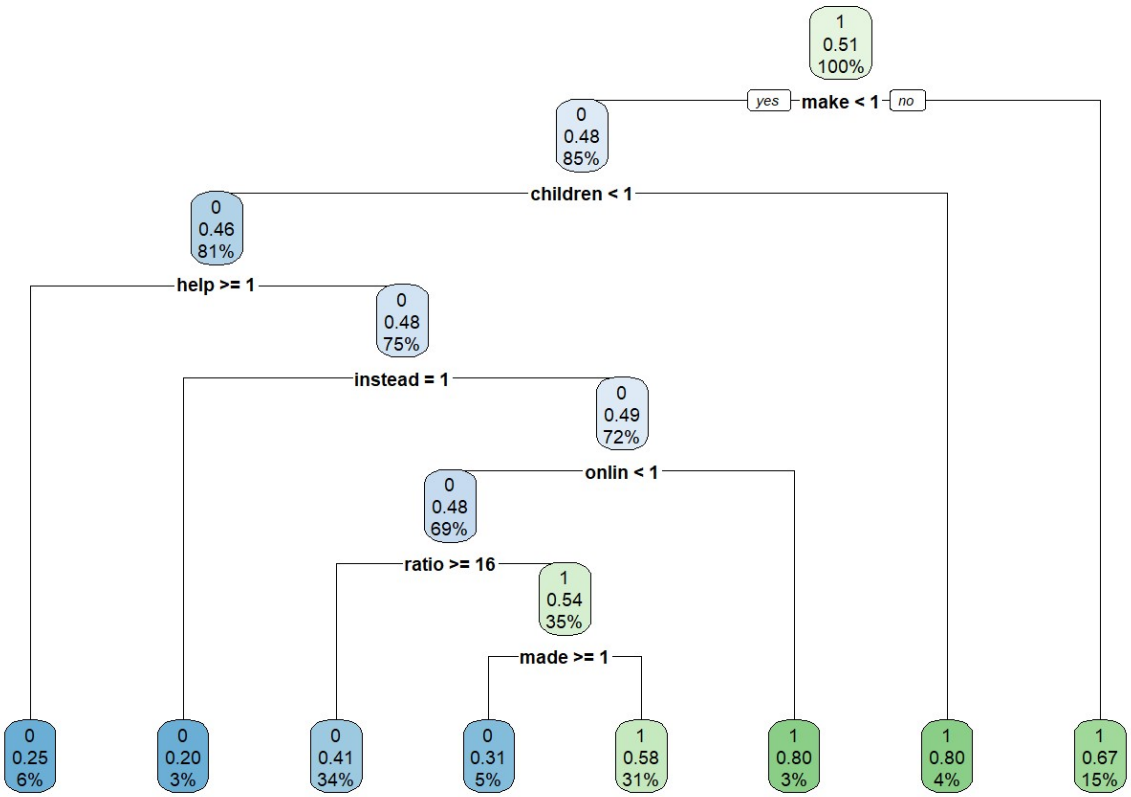
```
#Rebuild Model #

### Check the proportion of data #
nrow(subset(data_shark1, deal == 1))/nrow(data_shark1)
```

```
## [1] 0.5070707
```

```
#Partition the data into two parts Train and Test
set.seed(3000)
split = sample.split(data_shark1$deal, SplitRatio=0.7)
train_data1 = subset(data_shark1, split==T)
test_data1 = subset(data_shark1, split==F)
```

```
#CART#
library(rpart)
library(rpart.plot)
test_data1$deal = as.factor(test_data1$deal)
train_data1$deal = as.factor(train_data1$deal)
```

```
r.ctrl1 = rpart.control(minsplit = 100, minbucket = 10, cp = 0, xval = 10)
DTmodel1 = rpart(deal ~., data = train_data1, method = "class", control = r.ctrl1)
#CART Diagram
rpart.plot(DTmodel1)
```



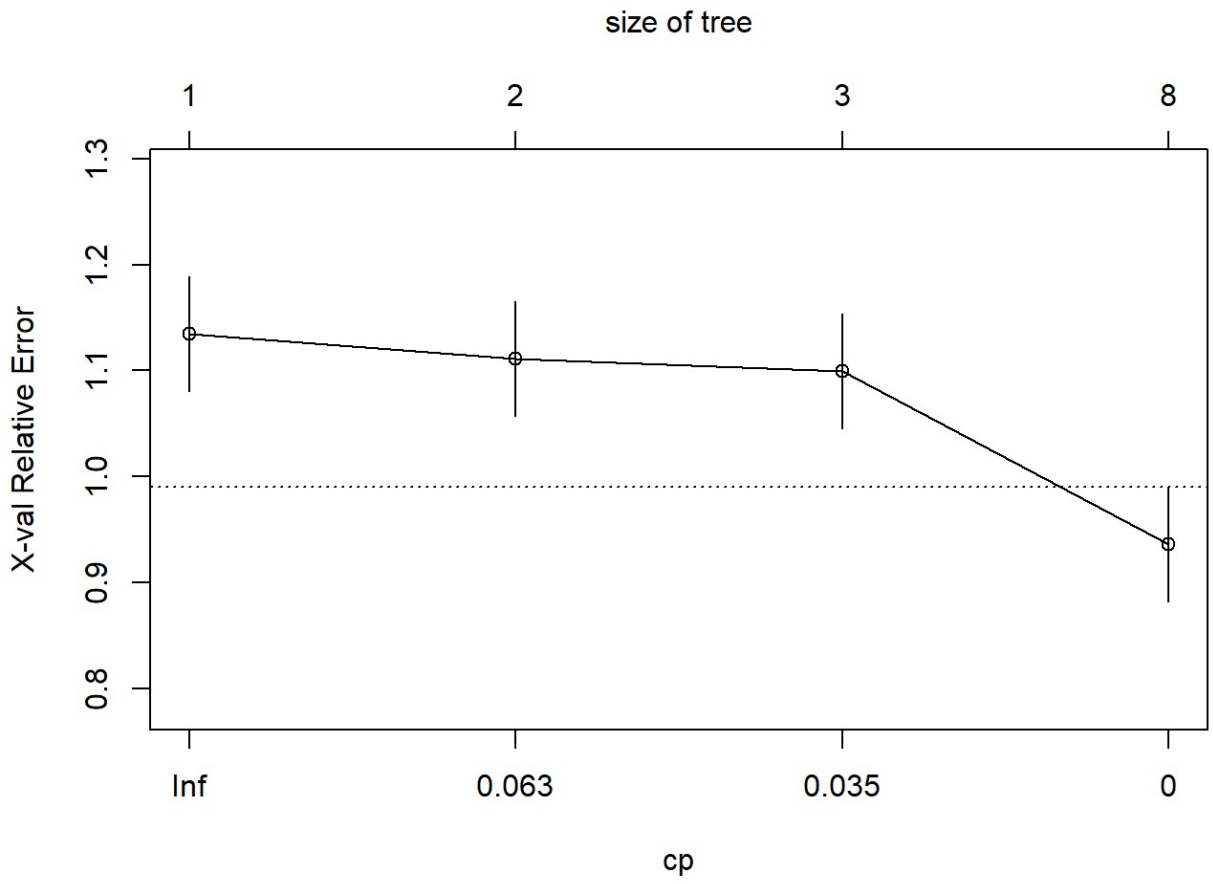

```
DTmodel1
```

```
## n= 347
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##   1) root 347 171 1 (0.4927954 0.5072046)
##     2) make< 0.5 295 141 0 (0.5220339 0.4779661)
##       4) children< 0.5 280 129 0 (0.5392857 0.4607143)
##         8) help>=0.5 20   5 0 (0.7500000 0.2500000) *
##         9) help< 0.5 260 124 0 (0.5230769 0.4769231)
##          18) instead>=0.5 10   2 0 (0.8000000 0.2000000) *
##          19) instead< 0.5 250 122 0 (0.5120000 0.4880000)
##            38) onlin< 0.5 240 114 0 (0.5250000 0.4750000)
##              76) ratio>=16 118  48 0 (0.5932203 0.4067797) *
##              77) ratio< 16 122  56 1 (0.4590164 0.5409836)
##               154) made>=0.5 16   5 0 (0.6875000 0.3125000) *
##               155) made< 0.5 106  45 1 (0.4245283 0.5754717) *
##            39) onlin>=0.5 10   2 1 (0.2000000 0.8000000) *
##       5) children>=0.5 15   3 1 (0.2000000 0.8000000) *
##     3) make>=0.5 52  17 1 (0.3269231 0.6730769) *
```

```
printcp(DTmodel1)
```

```
##
## Classification tree:
## rpart(formula = deal ~ ., data = train_data1, method = "class",
##     control = r.ctrl1)
##
## Variables actually used in tree construction:
## [1] children help     instead  made     make     onlin    ratio
##
## Root node error: 171/347 = 0.4928
##
## n= 347
##
##         CP nsplit rel error  xerror     xstd
## 1 0.076023      0   1.00000 1.13450 0.054086
## 2 0.052632      1   0.92398 1.11111 0.054221
## 3 0.023392      2   0.87135 1.09942 0.054277
## 4 0.000000      7   0.74269 0.93567 0.054303
```

```
plotcp(DTmodel1)
```



```
attributes(DTmodel1)
```

```
## $names
##  [1] "frame"              "where"              "call"
##  [4] "terms"              "cptable"            "method"
##  [7] "parms"              "control"            "functions"
## [10] "numresp"            "splits"             "variable.importance"
## [13] "y"                  "ordered"
##
## $xlevels
## named list()
##
## $ylevels
## [1] "0" "1"
##
## $class
## [1] "rpart"
```

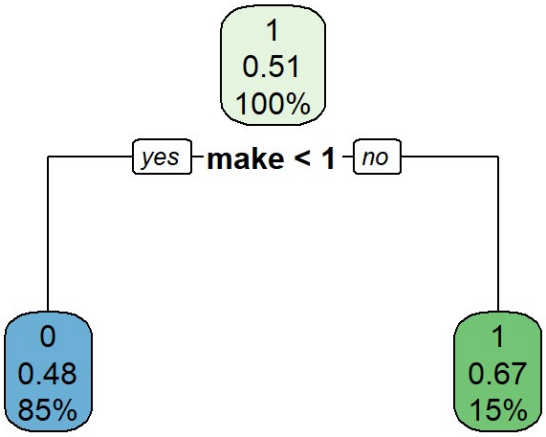
```
DTmodel1$cptable
```

```
##           CP nsplit rel error    xerror       xstd
## 1 0.07602339      0 1.0000000 1.1345029 0.05408611
## 2 0.05263158      1 0.9239766 1.1111111 0.05422078
## 3 0.02339181      2 0.8713450 1.0994152 0.05427710
## 4 0.00000000      7 0.7426901 0.9356725 0.05430252
```

```
ptree1 = prune(DTmodel1, 0.054, "CP")
print(ptree1)
```

```
## n= 347
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
## 1) root 347 171 1 (0.4927954 0.5072046)
##   2) make< 0.5 295 141 0 (0.5220339 0.4779661) *
##   3) make>=0.5 52  17 1 (0.3269231 0.6730769) *
```

```
rpart.plot(ptree1)
```




```
ptree1
```

```
## n= 347
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
## 1) root 347 171 1 (0.4927954 0.5072046)
##   2) make< 0.5 295 141 0 (0.5220339 0.4779661) *
##   3) make>=0.5 52  17 1 (0.3269231 0.6730769) *
```

```
DTmodel1$variable.importance
```

```
##          make      children         ratio         onlin          made
##    3.36580224    3.27784504    2.16068352    2.02800000    1.92271110
##          help       instead         order          book         innov
##    1.91263736    1.59507692    0.64593889    0.43704600    0.36050833
##         solut        easier          name         power    made_order
##    0.36050833    0.32363483    0.25890786    0.25890786    0.24033889
## natur_ingredi          care           mix         young        chocol
##    0.24033889    0.21852300    0.19418090    0.19418090    0.15950769
##       practic           fun          line        differ          find
##    0.15950769    0.12817614    0.12817614    0.09563187    0.09563187
##       kitchen          mani         might        design         enjoy
##    0.09563187    0.09563187    0.09563187    0.09155439    0.09155439
##         sport
##    0.09155439
```

```
#CART validation on test data
str(test_data1)
```

```
## 'data.frame':    148 obs. of  523 variables:
##  $ deal            : Factor w/ 2 levels "0","1": 1 1 2 1 2 1 2 2 2 2 ...
##  $ ratio           : int  14 10 15 20 35 10 15 20 25 10 ...
##  $ devic           : num  1 0 0 0 1 0 0 0 0 0 ...
##  $ retail          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ two             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ new             : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ parent          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ young           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ children        : num  0 0 0 0 0 0 2 0 0 0 ...
##  $ make            : num  0 0 0 0 0 0 1 1 0 0 ...
##  $ easi            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ littl           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ one             : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ turn            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ experi          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ play            : num  0 0 0 0 0 0 6 0 0 0 ...
##  $ provid          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ posit           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ organ           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pack            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ move            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ servic          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ deliv           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ colleg          : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ women           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ center          : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ room            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ offer           : num  0 1 0 0 0 0 1 0 0 0 ...
##  $ access          : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ educ            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ first           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ pitch           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ shark           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ tank            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ perfect         : num  0 0 2 0 0 0 0 0 0 0 ...
##  $ line            : num  0 0 1 1 0 0 0 0 0 0 ...
##  $ focus           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ gourmet         : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ food            : num  0 0 1 0 0 0 0 0 1 0 ...
##  $ product         : num  0 0 2 0 0 0 0 0 0 0 ...
##  $ sold            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ across          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ store           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ includ          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ dress           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ mani            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ design          : num  0 0 1 0 0 0 1 0 0 0 ...
```

```
##  $ flavor          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ shark_tank      : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ mix             : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ cloth           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ look            : num  0 0 0 1 0 0 1 0 0 0 ...
##  $ becom           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ big             : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ brand           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ activ           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ sport           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ apparel         : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ attach          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ hold            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ side            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ screen          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ safeti          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ prevent         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ start           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ vehicl          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ hous            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ get             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ student         : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ learn           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ classic         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ work            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ cook            : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ household       : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ item            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ made            : num  0 0 0 0 0 1 1 0 0 0 ...
##  $ recycl          : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ book            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ help            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ find            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ fit             : num  0 0 0 0 0 0 1 1 0 0 ...
##  $ fun             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ minut           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ effici          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ entertain       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ us              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ set             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ card            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ combin          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ keep            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ kid             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ long            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ school          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ develop         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ model           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ compani         : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ profession      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ remov           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ much            : num  0 0 0 0 0 0 2 0 0 0 ...
##   [list output truncated]
```

```
predCART2 = predict(ptree1, newdata = test_data1, type = "class")
predCART3 = predict(ptree1, newdata = test_data1, type = "prob")
predCART3
```

```
##             0         1
## 1   0.5220339 0.4779661
## 5   0.5220339 0.4779661
## 6   0.5220339 0.4779661
## 7   0.5220339 0.4779661
## 11  0.5220339 0.4779661
## 12  0.5220339 0.4779661
## 17  0.3269231 0.6730769
## 23  0.3269231 0.6730769
## 33  0.5220339 0.4779661
## 37  0.5220339 0.4779661
## 39  0.5220339 0.4779661
## 44  0.5220339 0.4779661
## 45  0.5220339 0.4779661
## 47  0.5220339 0.4779661
## 52  0.5220339 0.4779661
## 62  0.5220339 0.4779661
## 64  0.5220339 0.4779661
## 66  0.5220339 0.4779661
## 67  0.5220339 0.4779661
## 73  0.5220339 0.4779661
## 76  0.3269231 0.6730769
## 80  0.5220339 0.4779661
## 83  0.5220339 0.4779661
## 91  0.5220339 0.4779661
## 94  0.5220339 0.4779661
## 95  0.5220339 0.4779661
## 98  0.5220339 0.4779661
## 101 0.5220339 0.4779661
## 109 0.3269231 0.6730769
## 116 0.5220339 0.4779661
## 117 0.5220339 0.4779661
## 119 0.5220339 0.4779661
## 121 0.5220339 0.4779661
## 122 0.5220339 0.4779661
## 129 0.3269231 0.6730769
## 130 0.5220339 0.4779661
## 131 0.5220339 0.4779661
## 132 0.5220339 0.4779661
## 141 0.5220339 0.4779661
## 144 0.5220339 0.4779661
## 146 0.5220339 0.4779661
## 150 0.5220339 0.4779661
## 152 0.5220339 0.4779661
## 153 0.5220339 0.4779661
## 155 0.5220339 0.4779661
## 157 0.5220339 0.4779661
## 161 0.5220339 0.4779661
```

```
## 167 0.5220339 0.4779661
## 171 0.3269231 0.6730769
## 175 0.5220339 0.4779661
## 176 0.5220339 0.4779661
## 180 0.5220339 0.4779661
## 182 0.5220339 0.4779661
## 187 0.5220339 0.4779661
## 190 0.5220339 0.4779661
## 191 0.5220339 0.4779661
## 203 0.5220339 0.4779661
## 214 0.5220339 0.4779661
## 216 0.5220339 0.4779661
## 217 0.5220339 0.4779661
## 220 0.5220339 0.4779661
## 222 0.5220339 0.4779661
## 223 0.3269231 0.6730769
## 232 0.5220339 0.4779661
## 233 0.5220339 0.4779661
## 236 0.3269231 0.6730769
## 242 0.5220339 0.4779661
## 246 0.5220339 0.4779661
## 248 0.5220339 0.4779661
## 249 0.5220339 0.4779661
## 251 0.5220339 0.4779661
## 252 0.5220339 0.4779661
## 265 0.5220339 0.4779661
## 266 0.5220339 0.4779661
## 273 0.5220339 0.4779661
## 274 0.5220339 0.4779661
## 290 0.5220339 0.4779661
## 292 0.5220339 0.4779661
## 294 0.5220339 0.4779661
## 296 0.5220339 0.4779661
## 297 0.5220339 0.4779661
## 298 0.5220339 0.4779661
## 299 0.5220339 0.4779661
## 300 0.5220339 0.4779661
## 302 0.5220339 0.4779661
## 307 0.5220339 0.4779661
## 315 0.3269231 0.6730769
## 316 0.5220339 0.4779661
## 321 0.5220339 0.4779661
## 322 0.3269231 0.6730769
## 324 0.3269231 0.6730769
## 328 0.5220339 0.4779661
## 329 0.3269231 0.6730769
## 333 0.3269231 0.6730769
## 335 0.5220339 0.4779661
## 336 0.3269231 0.6730769
```

```
## 337 0.5220339 0.4779661
## 340 0.5220339 0.4779661
## 341 0.3269231 0.6730769
## 342 0.5220339 0.4779661
## 346 0.3269231 0.6730769
## 347 0.5220339 0.4779661
## 349 0.3269231 0.6730769
## 351 0.3269231 0.6730769
## 352 0.5220339 0.4779661
## 355 0.5220339 0.4779661
## 366 0.5220339 0.4779661
## 368 0.3269231 0.6730769
## 371 0.3269231 0.6730769
## 380 0.5220339 0.4779661
## 385 0.5220339 0.4779661
## 388 0.3269231 0.6730769
## 390 0.3269231 0.6730769
## 396 0.5220339 0.4779661
## 397 0.3269231 0.6730769
## 401 0.5220339 0.4779661
## 404 0.3269231 0.6730769
## 405 0.3269231 0.6730769
## 410 0.5220339 0.4779661
## 411 0.3269231 0.6730769
## 412 0.3269231 0.6730769
## 413 0.3269231 0.6730769
## 414 0.5220339 0.4779661
## 419 0.3269231 0.6730769
## 429 0.5220339 0.4779661
## 432 0.5220339 0.4779661
## 443 0.5220339 0.4779661
## 444 0.5220339 0.4779661
## 446 0.5220339 0.4779661
## 449 0.5220339 0.4779661
## 450 0.5220339 0.4779661
## 457 0.3269231 0.6730769
## 458 0.5220339 0.4779661
## 463 0.5220339 0.4779661
## 466 0.5220339 0.4779661
## 470 0.5220339 0.4779661
## 472 0.5220339 0.4779661
## 474 0.5220339 0.4779661
## 475 0.5220339 0.4779661
## 476 0.5220339 0.4779661
## 479 0.5220339 0.4779661
## 482 0.5220339 0.4779661
## 486 0.5220339 0.4779661
## 487 0.5220339 0.4779661
## 488 0.3269231 0.6730769
```

```
## 490 0.3269231 0.6730769
## 491 0.5220339 0.4779661
## 493 0.3269231 0.6730769
```

```
## deciling code
decile <- function(x){
  deciles <- vector(length=10)
  for (i in seq(0.1,1,.1)){
    deciles[i*10] <- quantile(x, i, na.rm=T)
  }
  return (
    ifelse(x<deciles[1], 1,
           ifelse(x<deciles[2], 2,
                  ifelse(x<deciles[3], 3,
                         ifelse(x<deciles[4], 4,
                                ifelse(x<deciles[5], 5,
                                       ifelse(x<deciles[6], 6,
                                              ifelse(x<deciles[7], 7,
                                                     ifelse(x<deciles[8], 8,
                                                            ifelse(x<deciles[9], 9, 10
                                                            ))))))))))
}


## deciling
test_data1$deciles <- decile(predCART3[,2])
test_data1$deal<- as.numeric(test_data1$deal)
str(test_data1)
```

```
## 'data.frame':    148 obs. of  524 variables:
##  $ deal            : num  1 1 2 1 2 1 2 2 2 2 ...
##  $ ratio           : int  14 10 15 20 35 10 15 20 25 10 ...
##  $ devic           : num  1 0 0 0 1 0 0 0 0 0 ...
##  $ retail          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ two             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ new             : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ parent          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ young           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ children        : num  0 0 0 0 0 0 2 0 0 0 ...
##  $ make            : num  0 0 0 0 0 0 1 1 0 0 ...
##  $ easi            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ littl           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ one             : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ turn            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ experi          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ play            : num  0 0 0 0 0 0 6 0 0 0 ...
##  $ provid          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ posit           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ organ           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pack            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ move            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ servic          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ deliv           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ colleg          : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ women           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ center          : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ room            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ offer           : num  0 1 0 0 0 0 1 0 0 0 ...
##  $ access          : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ educ            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ first           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ pitch           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ shark           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ tank            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ perfect         : num  0 0 2 0 0 0 0 0 0 0 ...
##  $ line            : num  0 0 1 1 0 0 0 0 0 0 ...
##  $ focus           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ gourmet         : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ food            : num  0 0 1 0 0 0 0 0 1 0 ...
##  $ product         : num  0 0 2 0 0 0 0 0 0 0 ...
##  $ sold            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ across          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ store           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ includ          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ dress           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ mani            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ design          : num  0 0 1 0 0 0 1 0 0 0 ...
```

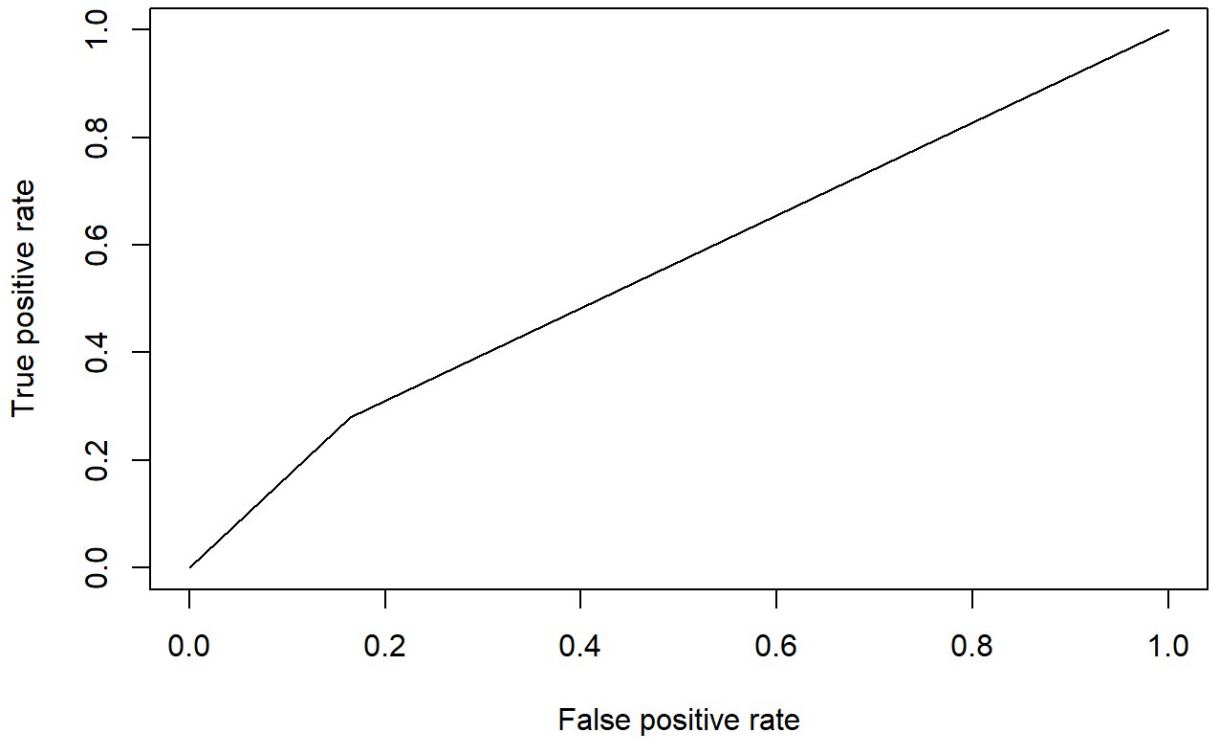
```
##  $ flavor          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ shark_tank      : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ mix             : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ cloth           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ look            : num  0 0 0 1 0 0 1 0 0 0 ...
##  $ becom           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ big             : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ brand           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ activ           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ sport           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ apparel         : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ attach          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ hold            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ side            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ screen          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ safeti          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ prevent         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ start           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ vehicl          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ hous            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ get             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ student         : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ learn           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ classic         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ work            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ cook            : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ household       : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ item            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ made            : num  0 0 0 0 0 1 1 0 0 0 ...
##  $ recycl          : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ book            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ help            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ find            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ fit             : num  0 0 0 0 0 0 1 1 0 0 ...
##  $ fun             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ minut           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ effici          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ entertain       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ us              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ set             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ card            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ combin          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ keep            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ kid             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ long            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ school          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ develop         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ model           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ compani         : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ profession      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ remov           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ much            : num  0 0 0 0 0 0 2 0 0 0 ...
##   [list output truncated]
```

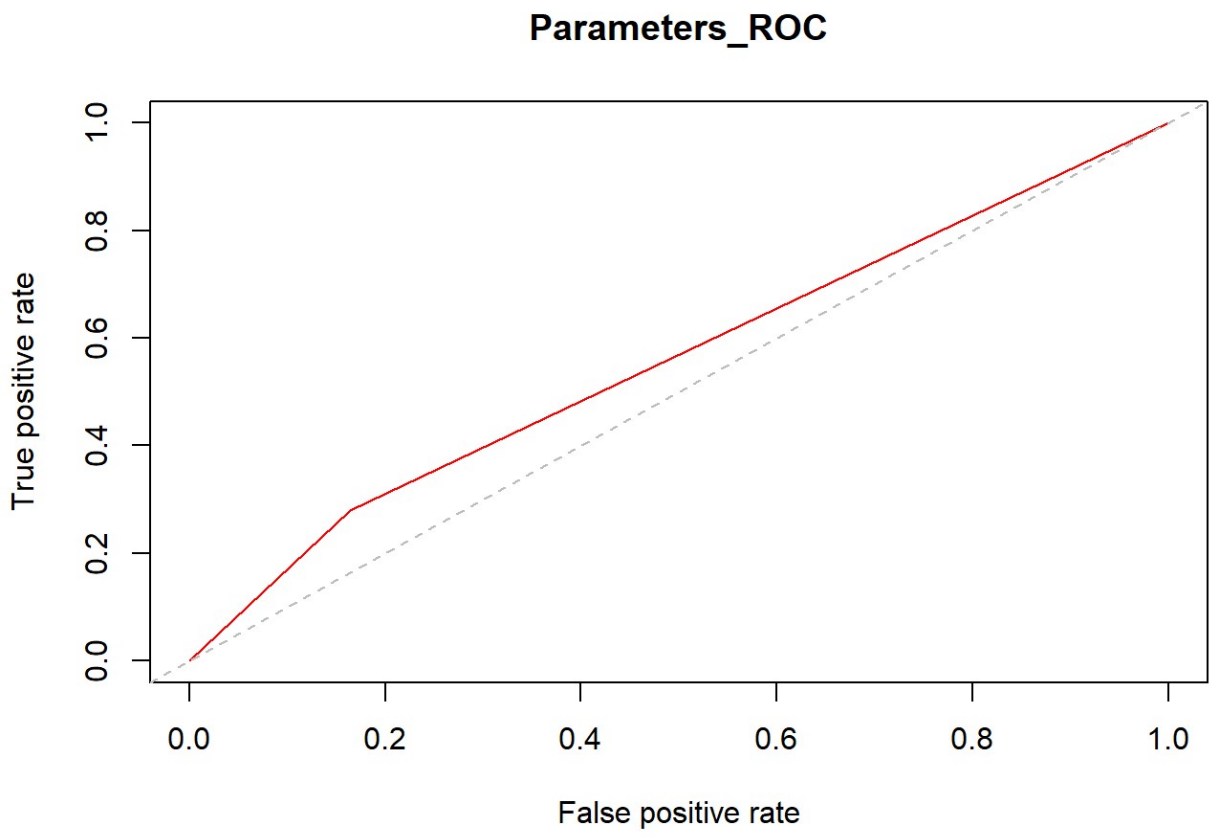
```
tmp_TS1 = data.table(test_data1)
n_rank <- tmp_TS1[, list(
  cnt = length(deal),
  cnt_resp = sum(deal),
  cnt_non_resp = sum(test_new$deal== 0)) ,
  by= deciles][order(- deciles)]
n_rank$rrate <- round (n_rank$cnt_resp / n_rank$cnt,2);
n_rank$cum_resp <- cumsum(n_rank$cnt_resp)
n_rank$cum_non_resp <- cumsum(n_rank$cnt_non_resp)
n_rank$cum_rel_resp <- round(n_rank$cum_resp / sum(n_rank$cnt_resp),2);
n_rank$cum_rel_non_resp <- round(n_rank$cum_non_resp / sum(n_rank$cnt_non_resp),2);
n_rank$ks <- abs(n_rank$cum_rel_resp - n_rank$cum_rel_non_resp);

library(scales)
n_rank$rrate <- percent(n_rank$rrate)
n_rank$cum_rel_resp <- percent(n_rank$cum_rel_resp)
n_rank$cum_rel_non_resp <- percent(n_rank$cum_rel_non_resp)

print(n_rank)
```

```
##    deciles cnt cnt_resp cnt_non_resp  rrate cum_resp cum_non_resp
## 1:      10  33       54           71 164.0%       54           71
## 2:       8 115      169           71 147.0%      223          142
##    cum_rel_resp cum_rel_non_resp   ks
## 1:        24.0%            50.0% 0.26
## 2:       100.0%           100.0% 0.00
```

```
#install.packages("ROCR")
library(ROCR)
#Validation on test data
DTpredROC1_new = ROCR::prediction(predCART3[,2], test_data1$deal)
perf1_new<-ROCR::performance(DTpredROC1_new,"tpr","fpr")
plot(perf1_new)
```




```
plot(perf1_new,col="red", main="Parameters_ROC")
abline(0,1, lty = 8, col = "grey")
```




```
##install.packages("ineq")
library(ineq)
#KS
KS <- max(attr(perf1_new, 'y.values')[[1]]-attr(perf1_new, 'x.values')[[1]])
KS
```

```
## [1] 0.1156164
```

```
auc <- as.numeric(ROCR::performance(DTpredROC1_new, "auc")@y.values)
auc
```

```
## [1] 0.5578082
```

```
#gini
gini = ineq(predCART3[,2], type="Gini")
gini
```

```
## [1] 0.06482461
```

```
with(test_data1,table(deal,predCART2))
```

```
##     predCART2
## deal  0  1
##    1 61 12
##    2 54 21
```

# Random forest

```
# Random forest#
str(train_data1)
```

```
## 'data.frame':    347 obs. of  523 variables:
##  $ deal            : Factor w/ 2 levels "0","1": 2 2 1 1 1 2 2 1 1 1 ...
##  $ ratio           : int  10 15 25 20 10 10 20 5 10 15 ...
##  $ devic           : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ retail          : num  2 0 0 0 0 0 0 0 0 0 ...
##  $ two             : num  1 0 0 0 0 0 0 0 0 0 ...
##  $ new             : num  1 0 0 0 0 0 0 0 0 0 ...
##  $ parent          : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ young           : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ children        : num  0 1 0 0 0 0 1 0 0 0 ...
##  $ make            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ easi            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ littl           : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ one             : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ turn            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ experi          : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ play            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ provid          : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ posit           : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ organ           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ pack            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ move            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ servic          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ deliv           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ colleg          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ women           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ center          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ room            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ offer           : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ access          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ educ            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ first           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pitch           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ shark           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ tank            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ perfect         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ line            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ focus           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gourmet         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ food            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ product         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sold            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ across          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ store           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ includ          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ dress           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ mani            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ design          : num  0 0 0 0 0 1 0 1 0 0 ...
```

```
##  $ flavor          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ shark_tank      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ mix             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ cloth           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ look            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ becom           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ big             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ brand           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ activ           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sport           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ apparel         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ attach          : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ hold            : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ side            : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ screen          : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ safeti          : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ prevent         : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ start           : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ vehicl          : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ hous            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ get             : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ student         : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ learn           : num  0 0 0 0 0 1 0 0 2 0 ...
##  $ classic         : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ work            : num  0 0 0 0 0 1 0 0 1 0 ...
##  $ cook            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ household       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ item            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ made            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ recycl          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ book            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ help            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ find            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ fit             : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ fun             : num  0 0 0 0 0 0 0 0 2 0 ...
##  $ minut           : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ effici          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ entertain       : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ us              : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ set             : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ card            : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ combin          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ keep            : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ kid             : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ long            : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ school          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ develop         : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ model           : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ compani         : num  0 0 0 0 0 0 0 0 0 1 ...
```

```
##  $ profession      : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ remov           : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ much            : num  0 0 0 0 0 0 0 0 0 0 ...
##   [list output truncated]
```

```
dim(train_data1)
```

```
## [1] 347 523
```

```
str(train_data1)
```

```
## 'data.frame':    347 obs. of  523 variables:
##  $ deal            : Factor w/ 2 levels "0","1": 2 2 1 1 1 2 2 1 1 1 ...
##  $ ratio           : int  10 15 25 20 10 10 20 5 10 15 ...
##  $ devic           : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ retail          : num  2 0 0 0 0 0 0 0 0 0 ...
##  $ two             : num  1 0 0 0 0 0 0 0 0 0 ...
##  $ new             : num  1 0 0 0 0 0 0 0 0 0 ...
##  $ parent          : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ young           : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ children        : num  0 1 0 0 0 0 1 0 0 0 ...
##  $ make            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ easi            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ littl           : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ one             : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ turn            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ experi          : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ play            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ provid          : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ posit           : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ organ           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ pack            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ move            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ servic          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ deliv           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ colleg          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ women           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ center          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ room            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ offer           : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ access          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ educ            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ first           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ pitch           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ shark           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ tank            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ perfect         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ line            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ focus           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ gourmet         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ food            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ product         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sold            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ across          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ store           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ includ          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ dress           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ mani            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ design          : num  0 0 0 0 0 1 0 1 0 0 ...
```

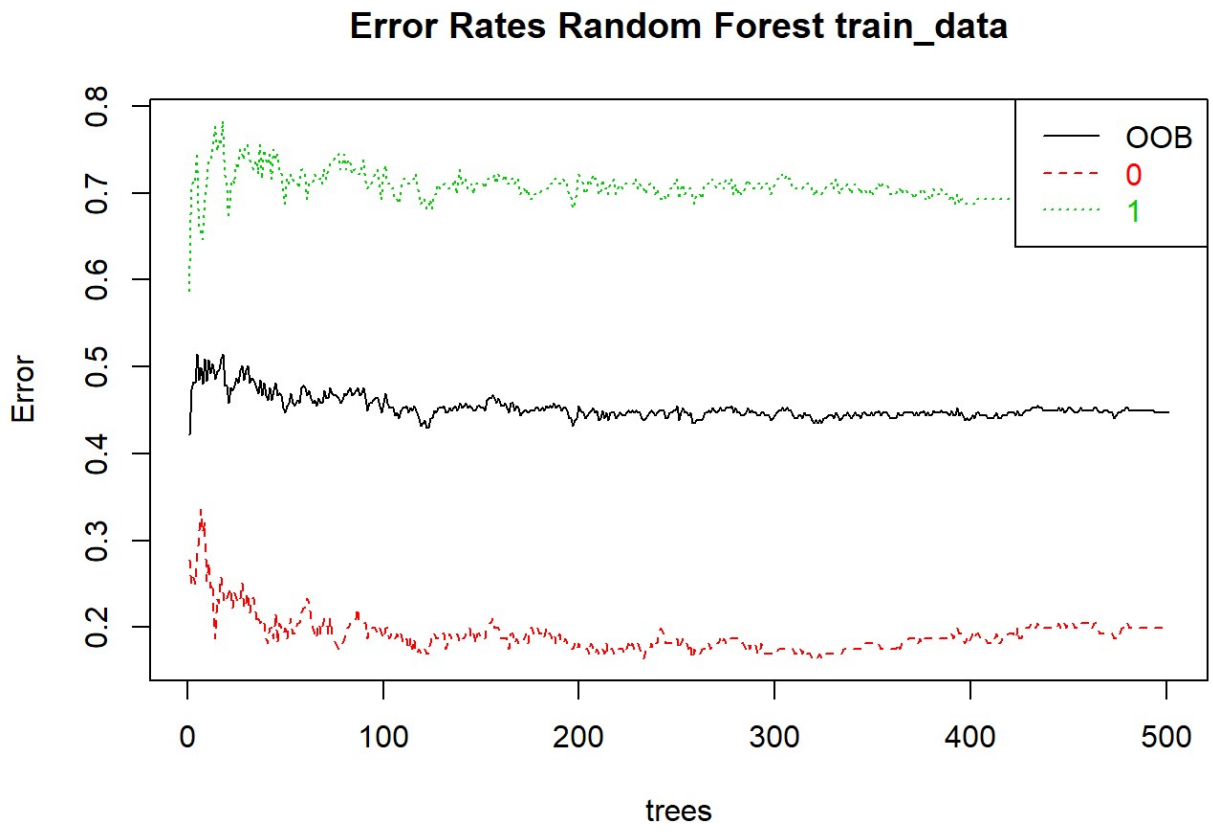
```
##  $ flavor          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ shark_tank      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ mix             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ cloth           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ look            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ becom           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ big             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ brand           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ activ           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sport           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ apparel         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ attach          : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ hold            : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ side            : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ screen          : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ safeti          : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ prevent         : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ start           : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ vehicl          : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ hous            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ get             : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ student         : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ learn           : num  0 0 0 0 0 1 0 0 2 0 ...
##  $ classic         : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ work            : num  0 0 0 0 0 1 0 0 1 0 ...
##  $ cook            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ household       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ item            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ made            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ recycl          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ book            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ help            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ find            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ fit             : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ fun             : num  0 0 0 0 0 0 0 0 2 0 ...
##  $ minut           : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ effici          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ entertain       : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ us              : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ set             : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ card            : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ combin          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ keep            : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ kid             : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ long            : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ school          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ develop         : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ model           : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ compani         : num  0 0 0 0 0 0 0 0 0 1 ...
```

```
##  $ profession      : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ remov           : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ much            : num  0 0 0 0 0 0 0 0 0 0 ...
##   [list output truncated]
```

```
colnames(train_data1) <- paste(colnames(train_data1), "_c", sep = "")
colnames(test_data1) <- paste(colnames(test_data1), "_c", sep = "")
colnames(train_data1)[1] <- 'deal'
colnames(test_data1)[1] <- 'deal'
train_data1$deal=as.factor(train_data1$deal)
library(randomForest)
seed=112
set.seed(seed)
RFmodel_new = randomForest(deal ~ ., data = train_data1, mtry = 3, nodesize = 7, ntree
= 501, importance = TRUE)
print(RFmodel_new)
```

```
##
## Call:
##  randomForest(formula = deal ~ ., data = train_data1, mtry = 3,      nodesize = 7,
ntree = 501, importance = TRUE)
##                Type of random forest: classification
##                      Number of trees: 501
## No. of variables tried at each split: 3
##
##         OOB estimate of  error rate: 44.67%
## Confusion matrix:
##     0  1 class.error
## 0 137 34   0.1988304
## 1 121 55   0.6875000
```

```
plot(RFmodel_new, main="")
legend("topright", c("OOB", "0", "1"), text.col=1:6, lty=1:3, col=1:3)
title(main="Error Rates Random Forest train_data")
```




```
RFmodel_new$err.rate
```

```
##              OOB         0         1
##   [1,] 0.4222222 0.2777778 0.5873016
##   [2,] 0.4741784 0.2500000 0.7047619
##   [3,] 0.4827586 0.2575758 0.7131783
##   [4,] 0.4809689 0.2500000 0.7103448
##   [5,] 0.5146580 0.2810458 0.7467532
##   [6,] 0.4845679 0.3062500 0.6585366
##   [7,] 0.4984985 0.3353659 0.6568047
##   [8,] 0.4805970 0.3090909 0.6470588
##   [9,] 0.5088235 0.3214286 0.6918605
##  [10,] 0.4839650 0.2529412 0.7109827
##  [11,] 0.5072464 0.2748538 0.7356322
##  [12,] 0.4927536 0.2456140 0.7356322
##  [13,] 0.5028902 0.2514620 0.7485714
##  [14,] 0.4855491 0.1871345 0.7771429
##  [15,] 0.4942197 0.2339181 0.7485714
##  [16,] 0.4971098 0.2280702 0.7600000
##  [17,] 0.5086705 0.2573099 0.7542857
##  [18,] 0.5144509 0.2397661 0.7828571
##  [19,] 0.4783862 0.2280702 0.7215909
##  [20,] 0.4783862 0.2339181 0.7159091
##  [21,] 0.4582133 0.2397661 0.6704545
##  [22,] 0.4755043 0.2456140 0.6988636
##  [23,] 0.4726225 0.2222222 0.7159091
##  [24,] 0.4783862 0.2397661 0.7102273
##  [25,] 0.4870317 0.2339181 0.7329545
##  [26,] 0.4812680 0.2280702 0.7272727
##  [27,] 0.4956772 0.2339181 0.7500000
##  [28,] 0.5014409 0.2514620 0.7443182
##  [29,] 0.4841499 0.2222222 0.7386364
##  [30,] 0.4956772 0.2280702 0.7556818
##  [31,] 0.5014409 0.2397661 0.7556818
##  [32,] 0.4812680 0.2163743 0.7386364
##  [33,] 0.4870317 0.2339181 0.7329545
##  [34,] 0.4841499 0.2339181 0.7272727
##  [35,] 0.4755043 0.2046784 0.7386364
##  [36,] 0.4697406 0.2105263 0.7215909
##  [37,] 0.4841499 0.2046784 0.7556818
##  [38,] 0.4668588 0.2105263 0.7159091
##  [39,] 0.4812680 0.2046784 0.7500000
##  [40,] 0.4668588 0.1871345 0.7386364
##  [41,] 0.4610951 0.1812865 0.7329545
##  [42,] 0.4755043 0.1988304 0.7443182
##  [43,] 0.4610951 0.1988304 0.7159091
##  [44,] 0.4726225 0.1871345 0.7500000
##  [45,] 0.4812680 0.2163743 0.7386364
##  [46,] 0.4668588 0.1812865 0.7443182
##  [47,] 0.4697406 0.2046784 0.7272727
```

```
##  [48,] 0.4668588 0.1988304 0.7272727
##  [49,] 0.4524496 0.1929825 0.7045455
##  [50,] 0.4466859 0.1988304 0.6875000
##  [51,] 0.4553314 0.1871345 0.7159091
##  [52,] 0.4582133 0.1988304 0.7102273
##  [53,] 0.4697406 0.2105263 0.7215909
##  [54,] 0.4582133 0.1929825 0.7159091
##  [55,] 0.4553314 0.1929825 0.7102273
##  [56,] 0.4610951 0.2046784 0.7102273
##  [57,] 0.4582133 0.2046784 0.7045455
##  [58,] 0.4755043 0.2163743 0.7272727
##  [59,] 0.4783862 0.2222222 0.7272727
##  [60,] 0.4755043 0.2222222 0.7215909
##  [61,] 0.4668588 0.2339181 0.6931818
##  [62,] 0.4726225 0.2280702 0.7102273
##  [63,] 0.4639769 0.2105263 0.7102273
##  [64,] 0.4582133 0.1988304 0.7102273
##  [65,] 0.4610951 0.1929825 0.7215909
##  [66,] 0.4553314 0.1871345 0.7159091
##  [67,] 0.4639769 0.2046784 0.7159091
##  [68,] 0.4582133 0.1929825 0.7159091
##  [69,] 0.4582133 0.1988304 0.7102273
##  [70,] 0.4726225 0.2105263 0.7272727
##  [71,] 0.4639769 0.2046784 0.7159091
##  [72,] 0.4639769 0.1929825 0.7272727
##  [73,] 0.4755043 0.2105263 0.7329545
##  [74,] 0.4697406 0.1929825 0.7386364
##  [75,] 0.4668588 0.1871345 0.7386364
##  [76,] 0.4668588 0.1812865 0.7443182
##  [77,] 0.4639769 0.1754386 0.7443182
##  [78,] 0.4582133 0.1812865 0.7272727
##  [79,] 0.4610951 0.1812865 0.7329545
##  [80,] 0.4697406 0.1871345 0.7443182
##  [81,] 0.4668588 0.1988304 0.7272727
##  [82,] 0.4726225 0.1988304 0.7386364
##  [83,] 0.4755043 0.2046784 0.7386364
##  [84,] 0.4668588 0.2046784 0.7215909
##  [85,] 0.4697406 0.2046784 0.7272727
##  [86,] 0.4726225 0.2105263 0.7272727
##  [87,] 0.4755043 0.2222222 0.7215909
##  [88,] 0.4668588 0.2046784 0.7215909
##  [89,] 0.4697406 0.2046784 0.7272727
##  [90,] 0.4755043 0.2046784 0.7386364
##  [91,] 0.4639769 0.2046784 0.7159091
##  [92,] 0.4495677 0.1871345 0.7045455
##  [93,] 0.4582133 0.1988304 0.7102273
##  [94,] 0.4582133 0.1988304 0.7102273
##  [95,] 0.4610951 0.1988304 0.7159091
##  [96,] 0.4639769 0.1988304 0.7215909
```

```
##  [97,] 0.4639769 0.1929825 0.7272727
##  [98,] 0.4553314 0.1988304 0.7045455
##  [99,] 0.4466859 0.1929825 0.6931818
## [100,] 0.4582133 0.1871345 0.7215909
## [101,] 0.4697406 0.1988304 0.7329545
## [102,] 0.4582133 0.1988304 0.7102273
## [103,] 0.4524496 0.1871345 0.7102273
## [104,] 0.4524496 0.1929825 0.7045455
## [105,] 0.4524496 0.1929825 0.7045455
## [106,] 0.4438040 0.1871345 0.6931818
## [107,] 0.4466859 0.1929825 0.6931818
## [108,] 0.4409222 0.1871345 0.6875000
## [109,] 0.4495677 0.1929825 0.6988636
## [110,] 0.4524496 0.1812865 0.7159091
## [111,] 0.4524496 0.1812865 0.7159091
## [112,] 0.4553314 0.1871345 0.7159091
## [113,] 0.4466859 0.1754386 0.7102273
## [114,] 0.4524496 0.1871345 0.7102273
## [115,] 0.4495677 0.1754386 0.7159091
## [116,] 0.4553314 0.1871345 0.7159091
## [117,] 0.4495677 0.1695906 0.7215909
## [118,] 0.4409222 0.1754386 0.6988636
## [119,] 0.4322767 0.1695906 0.6875000
## [120,] 0.4351585 0.1695906 0.6931818
## [121,] 0.4380403 0.1754386 0.6931818
## [122,] 0.4293948 0.1695906 0.6818182
## [123,] 0.4293948 0.1695906 0.6818182
## [124,] 0.4409222 0.1812865 0.6931818
## [125,] 0.4409222 0.1929825 0.6818182
## [126,] 0.4495677 0.1929825 0.6988636
## [127,] 0.4466859 0.1871345 0.6988636
## [128,] 0.4524496 0.1871345 0.7102273
## [129,] 0.4524496 0.1929825 0.7045455
## [130,] 0.4524496 0.1929825 0.7045455
## [131,] 0.4553314 0.1929825 0.7102273
## [132,] 0.4466859 0.1754386 0.7102273
## [133,] 0.4466859 0.1812865 0.7045455
## [134,] 0.4524496 0.1929825 0.7045455
## [135,] 0.4495677 0.1871345 0.7045455
## [136,] 0.4524496 0.1871345 0.7102273
## [137,] 0.4553314 0.1871345 0.7159091
## [138,] 0.4495677 0.1929825 0.6988636
## [139,] 0.4582133 0.1812865 0.7272727
## [140,] 0.4524496 0.1871345 0.7102273
## [141,] 0.4553314 0.1871345 0.7159091
## [142,] 0.4582133 0.1929825 0.7159091
## [143,] 0.4524496 0.1929825 0.7045455
## [144,] 0.4553314 0.1988304 0.7045455
## [145,] 0.4524496 0.1871345 0.7102273
```

```
## [146,] 0.4495677 0.1929825 0.6988636
## [147,] 0.4495677 0.1812865 0.7102273
## [148,] 0.4553314 0.1929825 0.7102273
## [149,] 0.4553314 0.1929825 0.7102273
## [150,] 0.4524496 0.1871345 0.7102273
## [151,] 0.4524496 0.1871345 0.7102273
## [152,] 0.4495677 0.1871345 0.7045455
## [153,] 0.4610951 0.1988304 0.7159091
## [154,] 0.4639769 0.2046784 0.7159091
## [155,] 0.4639769 0.2046784 0.7159091
## [156,] 0.4668588 0.2105263 0.7159091
## [157,] 0.4639769 0.1988304 0.7215909
## [158,] 0.4582133 0.1988304 0.7102273
## [159,] 0.4639769 0.1988304 0.7215909
## [160,] 0.4553314 0.1871345 0.7159091
## [161,] 0.4553314 0.1871345 0.7159091
## [162,] 0.4582133 0.1871345 0.7215909
## [163,] 0.4524496 0.1871345 0.7102273
## [164,] 0.4466859 0.1754386 0.7102273
## [165,] 0.4582133 0.1929825 0.7159091
## [166,] 0.4582133 0.1929825 0.7159091
## [167,] 0.4495677 0.1812865 0.7102273
## [168,] 0.4553314 0.1871345 0.7159091
## [169,] 0.4524496 0.1812865 0.7159091
## [170,] 0.4438040 0.1812865 0.6988636
## [171,] 0.4466859 0.1812865 0.7045455
## [172,] 0.4495677 0.1871345 0.7045455
## [173,] 0.4495677 0.1929825 0.6988636
## [174,] 0.4466859 0.1812865 0.7045455
## [175,] 0.4495677 0.1988304 0.6931818
## [176,] 0.4495677 0.1988304 0.6931818
## [177,] 0.4495677 0.1929825 0.6988636
## [178,] 0.4524496 0.1988304 0.6988636
## [179,] 0.4524496 0.1988304 0.6988636
## [180,] 0.4495677 0.1929825 0.6988636
## [181,] 0.4524496 0.1929825 0.7045455
## [182,] 0.4495677 0.1871345 0.7045455
## [183,] 0.4553314 0.1988304 0.7045455
## [184,] 0.4524496 0.1929825 0.7045455
## [185,] 0.4553314 0.1929825 0.7102273
## [186,] 0.4524496 0.1871345 0.7102273
## [187,] 0.4582133 0.1929825 0.7159091
## [188,] 0.4524496 0.1812865 0.7159091
## [189,] 0.4553314 0.1871345 0.7159091
## [190,] 0.4553314 0.1929825 0.7102273
## [191,] 0.4495677 0.1812865 0.7102273
## [192,] 0.4524496 0.1812865 0.7159091
## [193,] 0.4524496 0.1871345 0.7102273
## [194,] 0.4495677 0.1812865 0.7102273
```

```
## [195,] 0.4409222 0.1754386 0.6988636
## [196,] 0.4409222 0.1754386 0.6988636
## [197,] 0.4322767 0.1754386 0.6818182
## [198,] 0.4380403 0.1812865 0.6875000
## [199,] 0.4409222 0.1754386 0.6988636
## [200,] 0.4553314 0.1812865 0.7215909
## [201,] 0.4495677 0.1754386 0.7159091
## [202,] 0.4466859 0.1754386 0.7102273
## [203,] 0.4409222 0.1754386 0.6988636
## [204,] 0.4438040 0.1754386 0.7045455
## [205,] 0.4466859 0.1695906 0.7159091
## [206,] 0.4495677 0.1812865 0.7102273
## [207,] 0.4524496 0.1754386 0.7215909
## [208,] 0.4466859 0.1695906 0.7159091
## [209,] 0.4495677 0.1754386 0.7159091
## [210,] 0.4380403 0.1695906 0.6988636
## [211,] 0.4409222 0.1754386 0.6988636
## [212,] 0.4495677 0.1754386 0.7159091
## [213,] 0.4409222 0.1695906 0.7045455
## [214,] 0.4466859 0.1812865 0.7045455
## [215,] 0.4380403 0.1695906 0.6988636
## [216,] 0.4466859 0.1754386 0.7102273
## [217,] 0.4466859 0.1754386 0.7102273
## [218,] 0.4466859 0.1695906 0.7159091
## [219,] 0.4409222 0.1754386 0.6988636
## [220,] 0.4438040 0.1754386 0.7045455
## [221,] 0.4466859 0.1754386 0.7102273
## [222,] 0.4438040 0.1754386 0.7045455
## [223,] 0.4438040 0.1695906 0.7102273
## [224,] 0.4466859 0.1754386 0.7102273
## [225,] 0.4495677 0.1812865 0.7102273
## [226,] 0.4466859 0.1754386 0.7102273
## [227,] 0.4495677 0.1812865 0.7102273
## [228,] 0.4466859 0.1754386 0.7102273
## [229,] 0.4495677 0.1754386 0.7159091
## [230,] 0.4466859 0.1812865 0.7045455
## [231,] 0.4466859 0.1754386 0.7102273
## [232,] 0.4409222 0.1695906 0.7045455
## [233,] 0.4380403 0.1637427 0.7045455
## [234,] 0.4409222 0.1754386 0.6988636
## [235,] 0.4438040 0.1754386 0.7045455
## [236,] 0.4466859 0.1754386 0.7102273
## [237,] 0.4495677 0.1812865 0.7102273
## [238,] 0.4466859 0.1754386 0.7102273
## [239,] 0.4438040 0.1812865 0.6988636
## [240,] 0.4466859 0.1812865 0.7045455
## [241,] 0.4495677 0.1929825 0.6988636
## [242,] 0.4495677 0.1988304 0.6931818
## [243,] 0.4495677 0.1929825 0.6988636
```

```
## [244,] 0.4409222 0.1812865 0.6931818
## [245,] 0.4409222 0.1812865 0.6931818
## [246,] 0.4466859 0.1812865 0.7045455
## [247,] 0.4466859 0.1812865 0.7045455
## [248,] 0.4495677 0.1812865 0.7102273
## [249,] 0.4495677 0.1812865 0.7102273
## [250,] 0.4553314 0.1871345 0.7159091
## [251,] 0.4380403 0.1754386 0.6931818
## [252,] 0.4466859 0.1812865 0.7045455
## [253,] 0.4524496 0.1871345 0.7102273
## [254,] 0.4466859 0.1871345 0.6988636
## [255,] 0.4466859 0.1812865 0.7045455
## [256,] 0.4438040 0.1754386 0.7045455
## [257,] 0.4466859 0.1812865 0.7045455
## [258,] 0.4351585 0.1637427 0.6988636
## [259,] 0.4351585 0.1754386 0.6875000
## [260,] 0.4380403 0.1695906 0.6988636
## [261,] 0.4380403 0.1695906 0.6988636
## [262,] 0.4380403 0.1695906 0.6988636
## [263,] 0.4380403 0.1754386 0.6931818
## [264,] 0.4466859 0.1754386 0.7102273
## [265,] 0.4466859 0.1754386 0.7102273
## [266,] 0.4495677 0.1812865 0.7102273
## [267,] 0.4466859 0.1754386 0.7102273
## [268,] 0.4524496 0.1812865 0.7159091
## [269,] 0.4495677 0.1812865 0.7102273
## [270,] 0.4466859 0.1812865 0.7045455
## [271,] 0.4495677 0.1812865 0.7102273
## [272,] 0.4524496 0.1871345 0.7102273
## [273,] 0.4495677 0.1812865 0.7102273
## [274,] 0.4495677 0.1871345 0.7045455
## [275,] 0.4524496 0.1812865 0.7159091
## [276,] 0.4524496 0.1871345 0.7102273
## [277,] 0.4524496 0.1871345 0.7102273
## [278,] 0.4553314 0.1871345 0.7159091
## [279,] 0.4466859 0.1871345 0.6988636
## [280,] 0.4524496 0.1871345 0.7102273
## [281,] 0.4466859 0.1871345 0.6988636
## [282,] 0.4466859 0.1812865 0.7045455
## [283,] 0.4438040 0.1812865 0.6988636
## [284,] 0.4438040 0.1812865 0.6988636
## [285,] 0.4466859 0.1754386 0.7102273
## [286,] 0.4524496 0.1812865 0.7159091
## [287,] 0.4466859 0.1754386 0.7102273
## [288,] 0.4495677 0.1812865 0.7102273
## [289,] 0.4466859 0.1812865 0.7045455
## [290,] 0.4466859 0.1812865 0.7045455
## [291,] 0.4438040 0.1695906 0.7102273
## [292,] 0.4438040 0.1695906 0.7102273
```

```
## [293,] 0.4466859 0.1812865 0.7045455
## [294,] 0.4466859 0.1754386 0.7102273
## [295,] 0.4438040 0.1695906 0.7102273
## [296,] 0.4438040 0.1695906 0.7102273
## [297,] 0.4438040 0.1695906 0.7102273
## [298,] 0.4380403 0.1695906 0.6988636
## [299,] 0.4409222 0.1695906 0.7045455
## [300,] 0.4438040 0.1695906 0.7102273
## [301,] 0.4466859 0.1695906 0.7159091
## [302,] 0.4495677 0.1754386 0.7159091
## [303,] 0.4495677 0.1754386 0.7159091
## [304,] 0.4524496 0.1754386 0.7215909
## [305,] 0.4495677 0.1754386 0.7159091
## [306,] 0.4524496 0.1754386 0.7215909
## [307,] 0.4495677 0.1754386 0.7159091
## [308,] 0.4495677 0.1754386 0.7159091
## [309,] 0.4438040 0.1754386 0.7045455
## [310,] 0.4409222 0.1695906 0.7045455
## [311,] 0.4409222 0.1695906 0.7045455
## [312,] 0.4438040 0.1754386 0.7045455
## [313,] 0.4438040 0.1754386 0.7045455
## [314,] 0.4409222 0.1695906 0.7045455
## [315,] 0.4466859 0.1754386 0.7102273
## [316,] 0.4438040 0.1695906 0.7102273
## [317,] 0.4438040 0.1695906 0.7102273
## [318,] 0.4438040 0.1695906 0.7102273
## [319,] 0.4380403 0.1637427 0.7045455
## [320,] 0.4351585 0.1637427 0.6988636
## [321,] 0.4380403 0.1637427 0.7045455
## [322,] 0.4351585 0.1637427 0.6988636
## [323,] 0.4380403 0.1695906 0.6988636
## [324,] 0.4351585 0.1637427 0.6988636
## [325,] 0.4409222 0.1695906 0.7045455
## [326,] 0.4409222 0.1695906 0.7045455
## [327,] 0.4438040 0.1695906 0.7102273
## [328,] 0.4438040 0.1695906 0.7102273
## [329,] 0.4438040 0.1695906 0.7102273
## [330,] 0.4409222 0.1695906 0.7045455
## [331,] 0.4466859 0.1695906 0.7159091
## [332,] 0.4438040 0.1695906 0.7102273
## [333,] 0.4380403 0.1695906 0.6988636
## [334,] 0.4409222 0.1695906 0.7045455
## [335,] 0.4438040 0.1754386 0.7045455
## [336,] 0.4438040 0.1754386 0.7045455
## [337,] 0.4466859 0.1754386 0.7102273
## [338,] 0.4438040 0.1754386 0.7045455
## [339,] 0.4409222 0.1754386 0.6988636
## [340,] 0.4438040 0.1754386 0.7045455
## [341,] 0.4466859 0.1754386 0.7102273
```

```
## [342,] 0.4438040 0.1754386 0.7045455
## [343,] 0.4438040 0.1754386 0.7045455
## [344,] 0.4409222 0.1754386 0.6988636
## [345,] 0.4466859 0.1812865 0.7045455
## [346,] 0.4466859 0.1812865 0.7045455
## [347,] 0.4438040 0.1812865 0.6988636
## [348,] 0.4438040 0.1812865 0.6988636
## [349,] 0.4466859 0.1812865 0.7045455
## [350,] 0.4438040 0.1812865 0.6988636
## [351,] 0.4466859 0.1812865 0.7045455
## [352,] 0.4466859 0.1812865 0.7045455
## [353,] 0.4495677 0.1812865 0.7102273
## [354,] 0.4438040 0.1754386 0.7045455
## [355,] 0.4466859 0.1754386 0.7102273
## [356,] 0.4466859 0.1754386 0.7102273
## [357,] 0.4466859 0.1754386 0.7102273
## [358,] 0.4438040 0.1754386 0.7045455
## [359,] 0.4409222 0.1754386 0.6988636
## [360,] 0.4409222 0.1754386 0.6988636
## [361,] 0.4409222 0.1754386 0.6988636
## [362,] 0.4409222 0.1754386 0.6988636
## [363,] 0.4466859 0.1812865 0.7045455
## [364,] 0.4438040 0.1754386 0.7045455
## [365,] 0.4438040 0.1812865 0.6988636
## [366,] 0.4466859 0.1812865 0.7045455
## [367,] 0.4466859 0.1812865 0.7045455
## [368,] 0.4466859 0.1871345 0.6988636
## [369,] 0.4466859 0.1871345 0.6988636
## [370,] 0.4466859 0.1871345 0.6988636
## [371,] 0.4438040 0.1871345 0.6931818
## [372,] 0.4466859 0.1871345 0.6988636
## [373,] 0.4438040 0.1812865 0.6988636
## [374,] 0.4466859 0.1871345 0.6988636
## [375,] 0.4466859 0.1812865 0.7045455
## [376,] 0.4495677 0.1871345 0.7045455
## [377,] 0.4438040 0.1871345 0.6931818
## [378,] 0.4466859 0.1871345 0.6988636
## [379,] 0.4438040 0.1871345 0.6931818
## [380,] 0.4438040 0.1871345 0.6931818
## [381,] 0.4495677 0.1871345 0.7045455
## [382,] 0.4466859 0.1871345 0.6988636
## [383,] 0.4466859 0.1871345 0.6988636
## [384,] 0.4466859 0.1871345 0.6988636
## [385,] 0.4495677 0.1871345 0.7045455
## [386,] 0.4466859 0.1871345 0.6988636
## [387,] 0.4466859 0.1871345 0.6988636
## [388,] 0.4466859 0.1871345 0.6988636
## [389,] 0.4438040 0.1871345 0.6931818
## [390,] 0.4495677 0.1929825 0.6988636
```

```
## [391,] 0.4438040 0.1871345 0.6931818
## [392,] 0.4438040 0.1929825 0.6875000
## [393,] 0.4524496 0.1988304 0.6988636
## [394,] 0.4438040 0.1929825 0.6875000
## [395,] 0.4466859 0.1929825 0.6931818
## [396,] 0.4438040 0.1871345 0.6931818
## [397,] 0.4380403 0.1812865 0.6875000
## [398,] 0.4409222 0.1871345 0.6875000
## [399,] 0.4380403 0.1812865 0.6875000
## [400,] 0.4409222 0.1871345 0.6875000
## [401,] 0.4438040 0.1929825 0.6875000
## [402,] 0.4409222 0.1871345 0.6875000
## [403,] 0.4466859 0.1929825 0.6931818
## [404,] 0.4466859 0.1929825 0.6931818
## [405,] 0.4438040 0.1871345 0.6931818
## [406,] 0.4438040 0.1871345 0.6931818
## [407,] 0.4438040 0.1871345 0.6931818
## [408,] 0.4438040 0.1871345 0.6931818
## [409,] 0.4409222 0.1812865 0.6931818
## [410,] 0.4409222 0.1812865 0.6931818
## [411,] 0.4409222 0.1812865 0.6931818
## [412,] 0.4438040 0.1871345 0.6931818
## [413,] 0.4466859 0.1929825 0.6931818
## [414,] 0.4409222 0.1812865 0.6931818
## [415,] 0.4409222 0.1812865 0.6931818
## [416,] 0.4438040 0.1871345 0.6931818
## [417,] 0.4438040 0.1871345 0.6931818
## [418,] 0.4438040 0.1871345 0.6931818
## [419,] 0.4466859 0.1929825 0.6931818
## [420,] 0.4466859 0.1929825 0.6931818
## [421,] 0.4466859 0.1929825 0.6931818
## [422,] 0.4466859 0.1929825 0.6931818
## [423,] 0.4438040 0.1871345 0.6931818
## [424,] 0.4495677 0.1988304 0.6931818
## [425,] 0.4438040 0.1871345 0.6931818
## [426,] 0.4438040 0.1871345 0.6931818
## [427,] 0.4466859 0.1929825 0.6931818
## [428,] 0.4495677 0.1988304 0.6931818
## [429,] 0.4495677 0.1988304 0.6931818
## [430,] 0.4495677 0.1988304 0.6931818
## [431,] 0.4524496 0.2046784 0.6931818
## [432,] 0.4524496 0.2046784 0.6931818
## [433,] 0.4524496 0.1988304 0.6988636
## [434,] 0.4553314 0.2046784 0.6988636
## [435,] 0.4524496 0.1988304 0.6988636
## [436,] 0.4524496 0.1988304 0.6988636
## [437,] 0.4495677 0.1988304 0.6931818
## [438,] 0.4495677 0.1988304 0.6931818
## [439,] 0.4495677 0.1988304 0.6931818
```

```
## [440,] 0.4495677 0.1988304 0.6931818
## [441,] 0.4495677 0.1988304 0.6931818
## [442,] 0.4495677 0.1988304 0.6931818
## [443,] 0.4495677 0.1988304 0.6931818
## [444,] 0.4495677 0.1988304 0.6931818
## [445,] 0.4524496 0.2046784 0.6931818
## [446,] 0.4495677 0.1988304 0.6931818
## [447,] 0.4524496 0.2046784 0.6931818
## [448,] 0.4466859 0.2046784 0.6818182
## [449,] 0.4495677 0.1988304 0.6931818
## [450,] 0.4524496 0.2046784 0.6931818
## [451,] 0.4524496 0.2046784 0.6931818
## [452,] 0.4495677 0.2046784 0.6875000
## [453,] 0.4466859 0.1988304 0.6875000
## [454,] 0.4466859 0.1988304 0.6875000
## [455,] 0.4495677 0.1988304 0.6931818
## [456,] 0.4524496 0.2046784 0.6931818
## [457,] 0.4495677 0.2046784 0.6875000
## [458,] 0.4495677 0.2046784 0.6875000
## [459,] 0.4495677 0.2046784 0.6875000
## [460,] 0.4495677 0.2046784 0.6875000
## [461,] 0.4495677 0.2046784 0.6875000
## [462,] 0.4495677 0.1988304 0.6931818
## [463,] 0.4524496 0.2046784 0.6931818
## [464,] 0.4524496 0.2046784 0.6931818
## [465,] 0.4495677 0.1988304 0.6931818
## [466,] 0.4466859 0.1929825 0.6931818
## [467,] 0.4466859 0.1929825 0.6931818
## [468,] 0.4495677 0.1929825 0.6988636
## [469,] 0.4495677 0.1929825 0.6988636
## [470,] 0.4495677 0.1929825 0.6988636
## [471,] 0.4466859 0.1929825 0.6931818
## [472,] 0.4466859 0.1929825 0.6931818
## [473,] 0.4409222 0.1871345 0.6875000
## [474,] 0.4438040 0.1871345 0.6931818
## [475,] 0.4466859 0.1929825 0.6931818
## [476,] 0.4466859 0.1929825 0.6931818
## [477,] 0.4495677 0.1988304 0.6931818
## [478,] 0.4495677 0.1988304 0.6931818
## [479,] 0.4524496 0.2046784 0.6931818
## [480,] 0.4524496 0.2046784 0.6931818
## [481,] 0.4495677 0.1988304 0.6931818
## [482,] 0.4495677 0.1988304 0.6931818
## [483,] 0.4495677 0.1988304 0.6931818
## [484,] 0.4495677 0.1988304 0.6931818
## [485,] 0.4495677 0.1988304 0.6931818
## [486,] 0.4495677 0.1988304 0.6931818
## [487,] 0.4495677 0.1988304 0.6931818
## [488,] 0.4495677 0.1988304 0.6931818
```

```
## [489,] 0.4495677 0.1988304 0.6931818
## [490,] 0.4495677 0.1988304 0.6931818
## [491,] 0.4495677 0.1988304 0.6931818
## [492,] 0.4495677 0.1988304 0.6931818
## [493,] 0.4495677 0.1988304 0.6931818
## [494,] 0.4466859 0.1988304 0.6875000
## [495,] 0.4466859 0.1988304 0.6875000
## [496,] 0.4466859 0.1988304 0.6875000
## [497,] 0.4466859 0.1988304 0.6875000
## [498,] 0.4466859 0.1988304 0.6875000
## [499,] 0.4466859 0.1988304 0.6875000
## [500,] 0.4466859 0.1988304 0.6875000
## [501,] 0.4466859 0.1988304 0.6875000
```

```
rf_err_rate_new <- RFmodel_new$err.rate
rf_err_rate$ID <- seq.int(nrow(rf_err_rate_new))
```

```
## List the importance of the variables.
impVar1 <- round(randomForest::importance(RFmodel_new), 2)
impVar1[order(impVar1[,1],decreasing = TRUE),]
```

```
##                       0     1 MeanDecreaseAccuracy MeanDecreaseGini
## like_c             6.39  0.79                 4.97             0.48
## weight_c           5.85  2.70                 4.85             0.39
## make_c             5.61 -0.70                 4.03             0.50
## packag_c           5.52  2.18                 4.46             0.41
## shape_c            5.40  0.19                 3.96             0.31
## box_c              5.15  1.92                 4.34             0.41
## well_c             4.68 -0.99                 2.41             0.33
## super_c            4.67  0.83                 3.88             0.27
## season_c           4.63  3.40                 4.35             0.20
## patent_c           4.51  0.62                 3.21             0.40
## self_c             4.46  2.29                 3.74             0.24
## roll_c             4.44  3.76                 4.75             0.28
## water_c            4.42 -1.57                 2.84             0.24
## regular_c          4.34  1.17                 3.41             0.25
## name_c             4.31  1.74                 3.55             0.22
## clean_c            4.24  1.81                 3.61             0.27
## produc_c           4.24 -0.40                 2.40             0.20
## cream_c            4.04  4.20                 4.75             0.34
## young_c            4.03  1.34                 3.29             0.23
## children_c         4.02  3.50                 4.67             0.40
## connect_c          4.00  4.74                 4.77             0.21
## peanut_c           3.90  3.41                 4.10             0.13
## easi_c             3.82  0.82                 2.82             0.25
## free_c             3.78  2.51                 3.82             0.35
## can_c              3.72 -1.68                 1.93             0.25
## tradit_c           3.66  4.40                 4.49             0.35
## fire_c             3.64  2.56                 3.48             0.21
## turn_c             3.62  0.75                 2.89             0.21
## recycl_c           3.62  1.43                 2.79             0.20
## train_c            3.61  1.90                 3.03             0.24
## balanc_c           3.60  2.28                 3.35             0.17
## form_c             3.58  2.07                 3.39             0.37
## need_c             3.55 -2.14                 0.96             0.16
## safe_c             3.52 -1.77                 1.83             0.19
## solut_c            3.48 -0.43                 2.30             0.19
## easili_c           3.43 -0.97                 2.07             0.17
## ride_c             3.43  1.87                 3.14             0.20
## wheel_c            3.41  0.45                 2.18             0.21
## design_c           3.40  0.73                 2.62             0.37
## tri_c              3.37  0.49                 2.30             0.16
## great_c            3.31 -2.00                 1.20             0.16
## product_c          3.27 -1.34                 1.33             0.18
## work_c             3.23 -0.17                 1.81             0.26
## see_c              3.23  1.68                 3.22             0.13
## card_c             3.22  2.06                 3.20             0.24
## age_c              3.18 -1.79                 0.74             0.18
## less_c             3.14  1.46                 3.15             0.09
```

```
## app_c              3.13  1.35              2.57            0.19
## content_c          3.09  0.84              2.33            0.21
## put_c              3.08  1.08              2.63            0.07
## place_c            3.06 -1.16              1.16            0.13
## piec_c             3.05  0.01              1.87            0.14
## flavor_c           3.03 -1.70              1.22            0.29
## eas_c              3.01  2.66              3.43            0.15
## smart_c            3.00  0.22              2.01            0.18
## effect_c           2.99 -0.70              1.47            0.16
## butter_c           2.98  1.58              2.76            0.17
## open_c             2.96 -0.24              1.94            0.18
## men_c              2.92  2.73              3.69            0.31
## mani_c             2.90  1.81              3.18            0.14
## new_c              2.88  0.08              2.15            0.17
## littl_c            2.85 -1.05              0.93            0.09
## center_c           2.85  2.98              3.46            0.21
## might_c            2.83 -0.71              1.02            0.13
## even_c             2.82 -2.68              0.25            0.21
## yet_c              2.81 -0.81              1.37            0.22
## buy_c              2.81  0.10              1.85            0.13
## found_c            2.81 -2.05              0.23            0.16
## manufactur_c       2.76  0.84              2.17            0.18
## team_c             2.73  0.79              2.38            0.15
## contain_c          2.70 -2.85             -0.02            0.10
## get_c              2.66 -3.09             -0.39            0.19
## seat_c             2.66  2.01              2.84            0.11
## equip_c            2.66  2.08              3.03            0.20
## parent_c           2.64 -0.92              1.06            0.12
## small_c            2.63 -1.44              0.66            0.12
## beer_c             2.60  1.51              2.75            0.26
## stylish_c          2.59  1.10              2.14            0.18
## use_c              2.58 -2.21              0.28            0.19
## materi_c           2.54 -0.18              1.57            0.24
## store_c            2.53 -0.24              1.43            0.16
## space_c            2.53 -1.42              0.26            0.15
## item_c             2.46  0.41              1.78            0.08
## recip_c            2.44 -2.06              0.75            0.12
## educ_c             2.43  2.61              3.04            0.18
## test_c             2.43 -1.11              0.94            0.14
## bicycl_c           2.42  2.21              2.84            0.17
## serious_c          2.42 -1.39              0.62            0.06
## becom_c            2.41  1.45              2.26            0.14
## although_c         2.36  1.42              2.55            0.12
## stand_c            2.35 -1.42              1.91            0.05
## power_c            2.32 -1.18              0.58            0.16
## keep_c             2.31 -1.41              0.83            0.31
## cost_c             2.31  0.94              2.53            0.12
## togeth_c           2.30 -0.82              1.00            0.17
## natur_c            2.29 -3.18             -1.05            0.12
```

```
## typic_c          2.27  0.16              1.74          0.11
## car_c            2.25  0.98              1.97          0.12
## gear_c           2.24  0.42              1.39          0.08
## favorit_c        2.24  1.56              2.17          0.08
## alway_c          2.24  1.40              1.89          0.08
## promot_c         2.24 -0.99              0.61          0.13
## go_c             2.19 -1.91              0.37          0.09
## ratio_c          2.16  1.80              2.30          0.89
## award_c          2.14 -1.06              0.42          0.11
## ice_cream_c      2.14  2.84              2.96          0.20
## encourag_c       2.14 -0.92              0.91          0.14
## secur_c          2.13 -0.33              1.06          0.09
## tool_c           2.11 -1.83              0.24          0.11
## size_c           2.10 -1.86             -0.08          0.15
## feet_c           2.10  0.69              1.77          0.17
## thing_c          2.09  0.40              1.50          0.11
## blend_c          2.08 -1.53              0.54          0.15
## inch_c           2.03 -0.49              1.02          0.14
## save_c           2.02 -1.08              0.69          0.09
## adjust_c         2.02 -0.23              1.14          0.23
## sourc_c          2.01  0.81              1.74          0.08
## versatil_c       2.00 -0.31              1.38          0.12
## kid_c            1.99  0.10              1.38          0.15
## servic_c         1.98 -1.01              0.67          0.18
## brand_c          1.98 -0.90              0.55          0.18
## mother_c         1.98  1.42              2.15          0.11
## unlik_c          1.97 -2.00             -0.07          0.06
## non_c            1.97 -0.46              1.06          0.14
## companion_c      1.97  0.55              1.66          0.09
## five_c           1.95  0.21              1.36          0.12
## school_c         1.93 -0.91              0.87          0.10
## simpl_c          1.92 -0.51              0.84          0.15
## chocol_c         1.91  0.40              1.70          0.08
## win_c            1.86  0.82              1.51          0.12
## sport_c          1.85  1.20              1.74          0.18
## respons_c        1.85  1.73              2.03          0.08
## struggl_c        1.84 -0.06              1.35          0.08
## top_c            1.84 -0.84              0.67          0.10
## machin_c         1.83 -1.07              0.51          0.08
## mission_c        1.83  1.60              2.05          0.15
## effici_c         1.80 -0.26              0.80          0.10
## creat_c          1.80 -1.38              0.03          0.15
## choic_c          1.80  0.45              1.43          0.10
## resist_c         1.79  0.45              1.28          0.16
## retail_c         1.78  1.43              1.97          0.17
## behind_c         1.77  0.03              0.95          0.08
## just_c           1.76 -3.21             -1.07          0.19
## hand_c           1.76 -0.33              0.90          0.17
## caffein_c        1.75 -1.89             -0.01          0.07
```

```
## websit_c            1.74  0.37                1.40            0.08
## whole_c             1.73  0.79                1.73            0.13
## fast_c              1.73  2.52                2.88            0.14
## make_fashion_c      1.73  2.78                2.88            0.13
## goe_c               1.73 -0.83                0.25            0.09
## continu_c           1.67  0.07                1.16            0.05
## portion_c           1.66  0.66                1.30            0.09
## build_c             1.65 -2.17               -0.16            0.11
## easier_c            1.63 -0.39                0.82            0.10
## suit_c              1.63 -1.94               -0.89            0.06
## countri_c           1.61 -0.67                0.93            0.09
## interchang_c        1.61  2.77                2.73            0.29
## varieti_c           1.60 -1.33                0.06            0.16
## organ_c             1.59  0.03                0.89            0.12
## day_c               1.57  0.23                1.10            0.12
## call_c              1.57 -1.38                0.69            0.09
## tablet_c            1.56  2.24                2.37            0.14
## without_c           1.55 -1.88               -0.14            0.14
## social_c            1.55 -1.22                0.18            0.09
## may_c               1.55 -0.31                1.00            0.08
## damag_c             1.54 -0.66                1.03            0.05
## gift_c              1.52  0.68                1.18            0.18
## someth_c            1.52 -1.96               -0.01            0.06
## good_c              1.51 -2.18               -0.50            0.06
## sugar_c             1.51 -0.89                0.29            0.15
## air_c               1.51 -0.49                1.08            0.08
## develop_c           1.49 -2.45               -0.81            0.08
## ultim_c             1.49 -0.54                0.37            0.10
## pad_c               1.49 -1.88               -0.07            0.10
## shark_c             1.47 -2.34               -0.85            0.07
## bodi_c              1.47  1.99                2.09            0.35
## golf_c              1.47  1.23                1.91            0.12
## maintain_c          1.44 -1.36                0.05            0.06
## look_like_c         1.44 -2.14               -1.14            0.10
## smell_c             1.42 -1.64               -0.68            0.06
## display_c           1.42 -1.00                0.61            0.05
## start_c             1.41 -2.29               -0.52            0.12
## durabl_c            1.41 -0.83                0.43            0.09
## corpor_c            1.41  1.42                1.42            0.05
## learn_c             1.39 -1.06                0.41            0.11
## pair_c              1.39  2.04                2.11            0.19
## signatur_c          1.38 -1.60               -0.27            0.07
## event_c             1.38  1.22                1.63            0.13
## paper_c             1.37  0.23                0.92            0.09
## plan_c              1.34 -0.28                0.64            0.11
## expand_c            1.34 -2.91               -1.73            0.07
## accessori_c         1.33 -1.05                0.50            0.28
## print_c             1.32  0.05                0.84            0.13
## strong_c            1.31 -1.35               -0.52            0.09
```

```
## phone_c          1.29 -1.55            -0.11          0.09
## sustain_c        1.28  0.19             1.24          0.14
## magic_c          1.26  1.30             1.54          0.16
## women_c          1.25  0.49             1.16          0.14
## sold_c           1.24 -2.36            -0.50          0.13
## remov_c          1.20  2.04             1.91          0.18
## invent_c         1.20 -1.27            -0.36          0.10
## experi_c         1.19 -0.60             0.48          0.06
## larg_c           1.18  0.38             0.95          0.07
## provid_c         1.17 -0.28             0.71          0.10
## toy_c            1.16 -2.13            -0.58          0.12
## fulli_c          1.13 -0.88             0.16          0.05
## accord_c         1.13  0.23             0.91          0.10
## function_c       1.13 -0.23             0.58          0.14
## beach_c          1.12 -1.06            -0.19          0.11
## bed_c            1.10  0.48             1.06          0.17
## pay_c            1.10  0.36             0.94          0.06
## healthi_c        1.08 -2.41            -1.27          0.07
## stop_c           1.07 -1.41            -0.28          0.09
## ingredi_c        1.06 -0.50             0.38          0.12
## line_c           1.05 -1.78            -0.51          0.14
## light_c          1.05 -0.56             0.17          0.25
## sell_c           1.04 -2.11            -1.10          0.15
## user_c           1.03 -0.18             0.54          0.19
## cover_c          1.02 -0.25             0.29          0.17
## room_c           1.00  1.00             1.42          0.04
## throw_c          1.00  1.00             1.42          0.05
## direct_c         1.00  1.00             1.00          0.08
## stick_c          1.00 -0.66             0.27          0.06
## qualiti_c        1.00 -0.80             0.07          0.13
## conveni_c        1.00 -1.00            -0.03          0.06
## meal_c           1.00  1.04             1.29          0.10
## workout_c        1.00  0.00             1.00          0.08
## old_c            1.00 -1.11            -0.54          0.02
## bring_c          0.99 -1.38             0.26          0.05
## money_c          0.98 -0.40             0.41          0.07
## appli_c          0.97 -1.33            -0.38          0.10
## leg_c            0.96  1.00             1.14          0.08
## built_c          0.96  2.44             2.17          0.28
## fashion_c        0.95 -0.45             0.46          0.10
## gourmet_c        0.94  0.44             0.83          0.08
## now_c            0.94 -0.90             0.38          0.11
## individu_c       0.92 -0.70             0.07          0.07
## wed_c            0.89  0.49             0.89          0.05
## cut_c            0.86 -1.78            -0.05          0.10
## hair_c           0.86 -0.69             0.16          0.14
## compani_c        0.84  1.17             1.72          0.21
## dress_c          0.83  0.65             1.08          0.10
## consum_c         0.83 -2.12            -0.83          0.07
```

```
## month_c            0.81 -0.89              -0.05            0.07
## duo_c              0.77 -0.03               0.64            0.05
## three_c            0.76 -0.81              -0.23            0.12
## custom_c           0.74  2.41               1.99            0.23
## stuf_c             0.74 -1.09              -0.33            0.10
## storag_c           0.73  0.80               0.92            0.23
## ice_c              0.73  0.04               0.49            0.18
## expens_c           0.72 -1.71              -0.69            0.06
## everi_c            0.71 -0.07               0.41            0.14
## cloth_c            0.70 -1.18              -0.34            0.20
## color_c            0.69 -1.81              -0.78            0.18
## altern_c           0.68  2.30               1.83            0.20
## activ_c            0.67  0.22               0.77            0.09
## singl_c            0.66  0.55               0.75            0.12
## way_c              0.64 -2.00              -0.77            0.09
## anyon_c            0.64 -2.85              -1.64            0.09
## best_c             0.62 -1.93              -1.07            0.07
## sinc_c             0.60  0.62               0.69            0.03
## case_c             0.60 -0.37               0.07            0.08
## anoth_c            0.57 -0.02               0.43            0.08
## skin_c             0.57  2.14               1.65            0.20
## real_c             0.57  0.18               0.56            0.13
## help_c             0.56  0.20               0.48            0.28
## spray_c            0.56  0.72               0.65            0.12
## student_c          0.54  0.75               0.71            0.15
## featur_c           0.53 -1.42              -0.67            0.13
## inspir_c           0.52  0.37               0.53            0.07
## made_order_c       0.52 -0.74              -0.13            0.07
## first_c            0.51 -0.12               0.20            0.12
## much_c             0.50 -1.66              -0.80            0.04
## nut_c              0.49 -0.04               0.27            0.06
## wash_c             0.48 -1.57              -0.96            0.11
## rescu_c            0.47 -0.29              -0.05            0.05
## inflat_c           0.47 -1.08              -0.37            0.09
## made_usa_c         0.46 -1.45              -0.72            0.08
## innov_c            0.43 -1.59              -0.95            0.07
## oil_c              0.42 -1.02              -0.57            0.06
## window_c           0.41 -2.71              -1.48            0.12
## rather_c           0.41  0.86               0.85            0.10
## bike_c             0.40 -2.36              -1.37            0.05
## long_c             0.39 -1.93              -1.09            0.06
## chemic_c           0.38 -0.56              -0.01            0.07
## insid_c            0.37 -0.67              -0.29            0.11
## wast_c             0.36 -0.48              -0.02            0.10
## lightweight_c      0.35 -1.23              -0.76            0.09
## technolog_c        0.35 -1.29              -0.52            0.09
## allow_c            0.34 -1.59              -0.68            0.11
## avail_c            0.34 -1.63              -0.99            0.08
## chang_c            0.34 -1.07              -0.43            0.06
```

```
## california_c        0.33 -1.26              -0.56            0.07
## whether_c           0.32 -1.73              -0.90            0.08
## purchas_c           0.32 -1.96              -1.24            0.07
## glass_c             0.31  0.27               0.47            0.09
## process_c           0.29 -0.25               0.06            0.09
## price_c             0.27 -1.29              -0.59            0.10
## perfect_c           0.25 -1.22              -0.53            0.07
## eco_friend_c        0.25  0.83               0.53            0.06
## peopl_c             0.25 -1.68              -1.07            0.09
## coffe_c             0.24 -1.01              -0.76            0.10
## special_design_c    0.24 -0.92              -0.20            0.07
## choos_c             0.23 -0.38              -0.07            0.09
## layer_c             0.22 -0.31               0.00            0.09
## smartphon_c         0.22 -0.89              -0.52            0.07
## care_c              0.18 -2.31              -1.16            0.05
## theme_c             0.18 -1.04              -0.43            0.08
## order_c             0.16 -0.37              -0.19            0.09
## premium_c           0.16  1.05               0.76            0.09
## shark_tank_c        0.14 -0.49              -0.22            0.04
## babi_c              0.14  1.53               0.99            0.13
## person_c            0.14 -0.93              -0.59            0.11
## shirt_c             0.14 -1.69              -1.24            0.05
## simpli_c            0.14 -2.48              -1.83            0.06
## premium_qualiti_c   0.14 -1.86              -1.25            0.07
## sound_c             0.12 -0.71              -0.57            0.07
## screen_c            0.10  0.66               0.47            0.17
## cup_c               0.10  0.05               0.05            0.10
## natur_ingredi_c     0.09 -0.05               0.04            0.05
## safeti_c            0.08 -1.09              -0.53            0.13
## high_c              0.08  0.49               0.32            0.13
## toward_c            0.03 -2.30              -1.26            0.09
## afford_c            0.02 -1.86              -1.38            0.13
## tank_c              0.01 -0.89              -0.84            0.05
## home_c              0.01 -1.61              -1.36            0.15
## band_c              0.00  0.00               0.00            0.02
## lip_c               0.00 -1.00              -1.00            0.05
## instruct_c          0.00  0.00               0.00            0.02
## snack_c             0.00  0.00               0.00            0.05
## iphon_c             0.00  0.00               0.00            0.04
## deliv_c            -0.01  0.03               0.00            0.05
## stay_c             -0.01 -0.61              -0.40            0.03
## world_c            -0.01 -0.20              -0.16            0.17
## spend_c            -0.02  2.01               1.49            0.08
## pet_c              -0.02 -0.04               0.00            0.18
## system_c           -0.02 -1.50              -1.23            0.13
## right_c            -0.02 -2.24              -1.68            0.07
## ever_c             -0.03 -1.73              -0.87            0.04
## comfort_c          -0.03 -0.71              -0.37            0.08
## flagship_product_c -0.04  1.04               0.83            0.11
```

```
## kitchen_c          -0.06 -2.28              -1.56             0.09
## can_use_c          -0.06  0.66               0.40             0.04
## differ_c           -0.08 -0.76              -0.58             0.06
## video_c            -0.08 -2.32              -1.48             0.15
## support_c          -0.09 -2.01              -1.39             0.05
## outdoor_c          -0.09  0.71               0.51             0.06
## handl_c            -0.10 -1.42              -1.04             0.07
## show_c             -0.10 -0.05               0.02             0.08
## better_c           -0.11 -0.57              -0.37             0.08
## secret_c           -0.11 -2.56              -1.78             0.09
## bar_c              -0.15 -2.06              -1.18             0.11
## grow_c             -0.16 -0.69              -0.50             0.07
## tast_c             -0.16  0.54               0.39             0.12
## prevent_c          -0.17 -0.96              -0.68             0.12
## portabl_c          -0.17 -1.34              -0.97             0.07
## rang_c             -0.17  0.19              -0.04             0.10
## meet_c             -0.17 -1.94              -1.46             0.06
## usa_c              -0.18 -1.28              -1.04             0.07
## serv_c             -0.20  0.20              -0.01             0.10
## travel_c           -0.21 -0.31              -0.33             0.18
## partner_c          -0.21 -0.52              -0.47             0.08
## access_c           -0.25 -2.21              -1.56             0.07
## belt_c             -0.26 -1.82              -1.30             0.10
## pack_c             -0.27  1.03               0.34             0.16
## kind_c             -0.27 -0.66              -0.58             0.08
## kit_c              -0.27 -2.39              -1.82             0.11
## look_c             -0.28 -1.77              -1.32             0.12
## busi_c             -0.28 -2.20              -1.35             0.12
## music_c            -0.28 -2.65              -1.89             0.10
## thank_c            -0.29 -1.68              -1.25             0.06
## attract_c          -0.32  0.82               0.34             0.14
## clip_c             -0.33 -1.42              -0.79             0.04
## hous_c             -0.35 -0.45              -0.40             0.11
## audio_c            -0.36 -1.73              -1.19             0.04
## move_c             -0.38 -0.28              -0.57             0.10
## cours_c            -0.38 -0.05              -0.24             0.15
## feel_c             -0.39 -1.76              -1.29             0.05
## high_qualiti_c     -0.41 -1.77              -1.70             0.04
## flagship_c         -0.43 -2.20              -1.63             0.09
## preserv_c          -0.43 -0.75              -0.66             0.11
## cool_c             -0.43 -0.04              -0.24             0.12
## magnet_c           -0.45 -1.77              -1.23             0.11
## sturdi_c           -0.46 -0.52              -0.58             0.04
## delici_c           -0.46 -0.71              -0.72             0.06
## bright_c           -0.48 -1.00              -0.83             0.05
## complet_c          -0.48 -1.22              -1.23             0.08
## energi_c           -0.50 -2.66              -2.32             0.05
## gluten_c           -0.50 -0.45              -0.62             0.03
## ad_c               -0.52 -0.60              -0.84             0.11
```

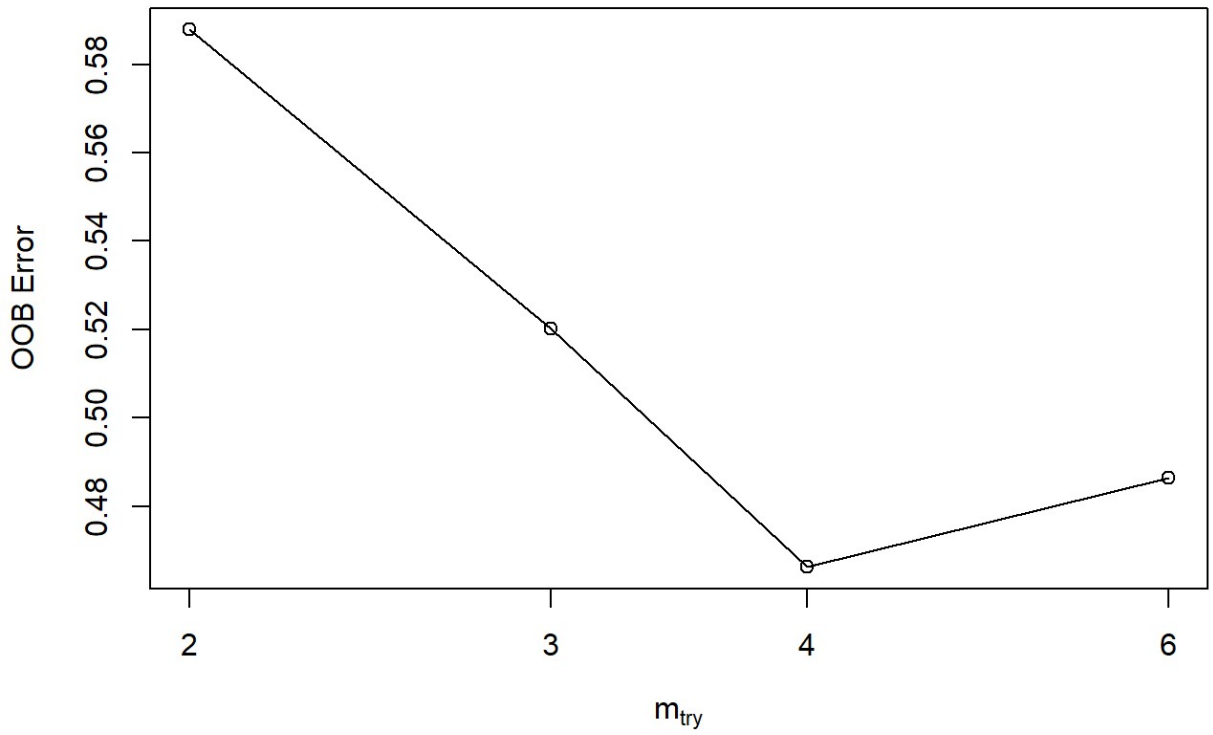
```
## onlin_c           -0.53 -1.15              -1.13            0.20
## low_c             -0.55 -1.03              -1.05            0.09
## select_c          -0.56 -1.11              -1.04            0.07
## instead_c         -0.58  1.51               0.59            0.32
## come_c            -0.59 -2.62              -2.09            0.17
## surfac_c          -0.60  0.01              -0.51            0.06
## pocket_c          -0.60  1.44               0.59            0.20
## mobil_c           -0.60 -2.19              -1.97            0.06
## classic_c         -0.61 -0.77              -1.00            0.04
## nation_c          -0.61  0.82               0.17            0.08
## wood_c            -0.62 -1.22              -1.33            0.06
## mom_c             -0.63  0.51              -0.27            0.07
## style_c           -0.65 -1.41              -1.31            0.17
## base_c            -0.66 -2.27              -1.87            0.14
## engin_c           -0.66 -0.67              -0.79            0.08
## one_c             -0.68 -3.01              -2.77            0.19
## let_c             -0.68 -0.76              -0.89            0.06
## transform_c       -0.69  1.18               0.21            0.09
## colleg_c          -0.70 -2.11              -1.85            0.10
## hold_c            -0.70 -2.04              -1.59            0.08
## market_c          -0.71 -2.52              -1.80            0.11
## set_c             -0.73 -1.17              -1.41            0.10
## us_c              -0.74  1.65               0.46            0.05
## motorcycl_c       -0.75 -1.67              -1.45            0.09
## fit_c             -0.76 -2.04              -1.54            0.10
## still_c           -0.76  0.83              -0.01            0.07
## time_c            -0.77 -2.39              -2.20            0.16
## year_c            -0.78 -1.07              -1.20            0.11
## devic_c           -0.80  0.54              -0.07            0.20
## take_c            -0.80  1.37               0.31            0.21
## plastic_c         -0.81 -2.30              -2.14            0.11
## readi_c           -0.81 -0.43              -0.87            0.09
## current_c         -0.81 -0.30              -0.61            0.06
## parti_c           -0.82 -1.60              -1.49            0.07
## game_c            -0.82 -1.09              -1.24            0.15
## famili_c          -0.83 -1.27              -1.39            0.09
## protect_c         -0.86 -2.37              -2.15            0.11
## tie_c             -0.86 -1.54              -1.68            0.08
## drink_c           -0.86 -0.74              -1.22            0.15
## posit_c           -0.90 -0.04              -0.64            0.06
## mix_c             -0.90 -3.15              -2.83            0.12
## shoe_c            -0.92  1.77               0.43            0.19
## full_c            -0.93 -1.61              -1.52            0.13
## ship_c            -0.93 -1.37              -1.54            0.10
## allow_user_c      -0.93 -0.61              -0.96            0.06
## combin_c          -0.94 -1.93              -2.20            0.08
## give_c            -0.94 -2.07              -1.80            0.11
## addit_c           -1.00 -1.42              -1.74            0.03
## improv_c          -1.00 -1.00              -1.42            0.02
```

```
## visibl_c          -1.01  0.85              -0.17            0.08
## apparel_c         -1.02 -2.09              -2.14            0.09
## fun_c             -1.02 -1.06              -1.28            0.18
## receiv_c          -1.04 -1.14              -1.52            0.06
## across_c          -1.05 -2.46              -2.15            0.03
## includ_c          -1.06 -2.10              -2.05            0.13
## detail_c          -1.06 -2.01              -1.74            0.03
## offer_c           -1.08 -2.09              -1.83            0.11
## flat_c            -1.09 -0.59              -0.93            0.05
## book_c            -1.12 -0.83              -1.17            0.07
## own_c             -1.14 -1.67              -1.48            0.04
## anim_c            -1.15 -1.32              -1.54            0.07
## focus_c           -1.17  0.44              -0.35            0.07
## sale_c            -1.17  1.73              -0.04            0.05
## point_c           -1.18 -1.06              -1.44            0.09
## two_c             -1.20  0.28              -0.68            0.11
## dog_c             -1.22 -0.71              -1.30            0.17
## practic_c         -1.23 -2.32              -2.11            0.05
## among_c           -1.24 -1.32              -1.45            0.09
## entertain_c       -1.26  0.00              -0.83            0.10
## number_c          -1.28 -2.16              -2.30            0.07
## origin_c          -1.30 -0.83              -1.41            0.04
## reusabl_c         -1.30 -0.55              -1.30            0.07
## add_c             -1.35 -2.46              -2.52            0.06
## heat_c            -1.37 -1.41              -1.63            0.06
## pitch_c           -1.42 -0.55              -1.01            0.08
## success_c         -1.42 -0.96              -1.25            0.08
## result_c          -1.42 -0.54              -0.94            0.10
## want_c            -1.44 -1.59              -1.79            0.09
## pre_c             -1.45 -1.81              -1.82            0.09
## rest_c            -1.45 -0.08              -1.18            0.06
## subscript_c       -1.46 -0.97              -1.36            0.07
## danc_c            -1.47 -1.43              -1.81            0.06
## minut_c           -1.48 -2.05              -2.11            0.10
## can_buy_c         -1.52 -0.77              -1.44            0.08
## quick_c           -1.55 -1.72              -1.87            0.13
## bottl_c           -1.56 -1.91              -2.38            0.10
## made_c            -1.60 -0.94              -1.39            0.27
## big_c             -1.61 -0.39              -1.35            0.06
## vehicl_c          -1.62 -1.55              -1.74            0.08
## program_c         -1.65 -0.78              -1.45            0.06
## kid_can_c         -1.73  0.66              -0.70            0.08
## wearer_c          -1.77 -1.80              -1.93            0.09
## find_c            -1.81  0.02              -1.30            0.10
## beverag_c         -1.81 -2.57              -2.59            0.11
## back_c            -1.82 -0.28              -1.54            0.10
## around_c          -1.84 -2.77              -2.44            0.10
## special_c         -1.88 -2.10              -2.38            0.09
## attach_c          -1.89 -1.21              -1.81            0.08
```

```
## household_c        -1.90 -1.59               -2.23            0.09
## friend_c           -1.91 -1.25               -1.76            0.15
## fresh_c            -1.93 -1.69               -1.98            0.07
## model_c            -1.98 -0.92               -1.79            0.08
## love_c             -1.99 -2.24               -2.35            0.06
## replac_c           -2.00 -2.24               -3.02            0.03
## enjoy_c            -2.01 -3.06               -2.88            0.12
## ball_c             -2.01 -0.45               -1.44            0.08
## prospect_c         -2.04 -1.31               -1.95            0.06
## treat_c            -2.05 -1.53               -2.24            0.12
## run_c              -2.06 -3.05               -2.98            0.09
## wine_c             -2.07 -0.01               -1.05            0.14
## carri_c            -2.10 -2.30               -2.82            0.06
## shop_c             -2.14 -2.59               -2.80            0.09
## live_c             -2.17 -3.63               -3.54            0.09
## eco_c              -2.18 -1.33               -2.04            0.10
## cake_c             -2.18 -1.95               -2.23            0.08
## control_c          -2.18 -1.67               -2.50            0.09
## uniqu_c            -2.26 -3.04               -3.05            0.08
## side_c             -2.34 -1.56               -2.45            0.07
## face_c             -2.35 -1.04               -2.41            0.09
## truck_c            -2.43 -2.50               -2.61            0.07
## chain_c            -2.48 -2.94               -3.16            0.09
## profession_c       -2.49 -1.67               -2.28            0.09
## bag_c              -2.57 -1.71               -2.57            0.12
## craft_c            -2.58 -1.96               -2.73            0.05
## cook_c             -2.59 -1.43               -2.42            0.05
## food_c             -2.64 -1.65               -2.64            0.11
## creativ_c          -2.65 -2.12               -2.81            0.09
## reduc_c            -2.72 -3.18               -3.46            0.08
## play_c             -2.89 -2.22               -2.84            0.06
## electr_c           -3.09 -0.41               -2.15            0.10
## remot_c            -3.12 -0.64               -2.51            0.10
## ultra_c            -3.19 -0.91               -3.03            0.06
```

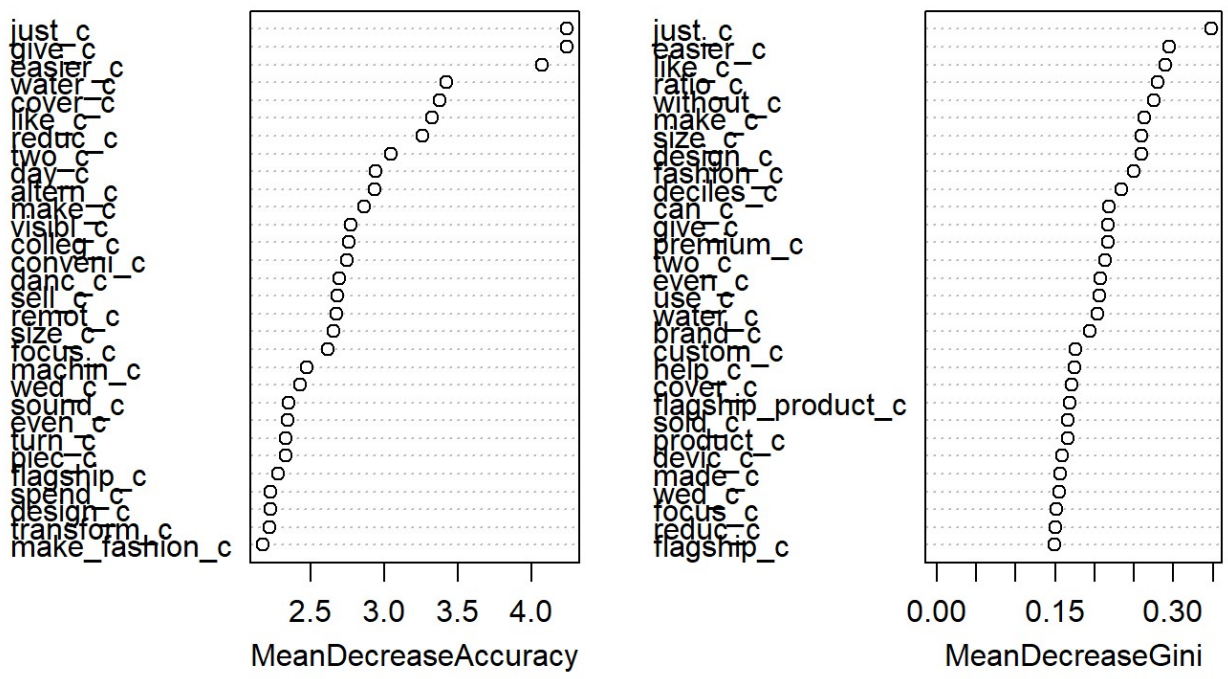
```
tune_rf_model_new <- tuneRF(x = test_data1[,-c(1)],
                        y=as.factor(test_data1$deal),
                        mtryStart = 3,
                        ntreeTry=100,
                        stepFactor = 1.5,
                        improve = 0.0001,
                        trace=TRUE,
                        plot = TRUE,
                        doBest = TRUE,
                        nodesize = 10,
                        importance=TRUE
)
```

```
## mtry = 3  OOB error = 52.03%
## Searching left ...
## mtry = 2     OOB error = 58.78%
## -0.1298701 1e-04
## Searching right ...
## mtry = 4     OOB error = 46.62%
## 0.1038961 1e-04
## mtry = 6     OOB error = 48.65%
## -0.04347826 1e-04
```




```
#Validate RF model on test data
test_data1$predict.class <- predict(tune_rf_model_new, test_data1, type="class")
test_data1$predict.score <- predict(tune_rf_model_new, test_data1, type="prob")
```

```
#Checking Variable Importance
varImpPlot(tune_rf_model_new)
```

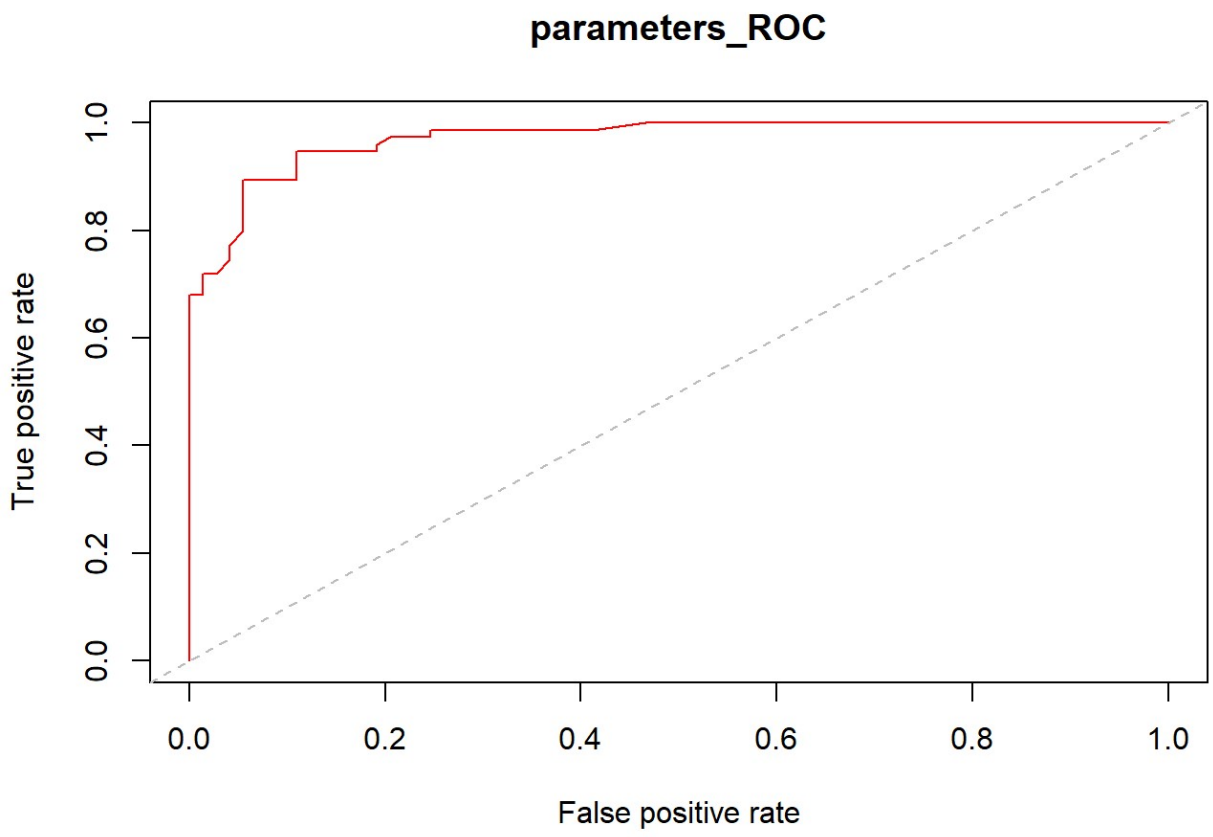



```
# deciling
decile <- function(x){
  deciles <- vector(length=10)
  for (i in seq(0.1,1,.1)){
    deciles[i*10] <- quantile(x, i, na.rm=T)
  }
  return (
    ifelse(x<deciles[1], 1,
           ifelse(x<deciles[2], 2,
                  ifelse(x<deciles[3], 3,
                         ifelse(x<deciles[4], 4,
                                ifelse(x<deciles[5], 5,
                                       ifelse(x<deciles[6], 6,
                                              ifelse(x<deciles[7], 7,
                                                     ifelse(x<deciles[8], 8,
                                                            ifelse(x<deciles[9], 9, 10
                                                            ))))))))))
}


test_data1$deciles <- decile(test_data1$predict.score[,2])
class(test_data1$predict.score)
```

```
## [1] "matrix" "votes"
```

```
library(tidyverse)
library(magrittr)
#train_data1[, 1] <- as.numeric(as.character( train_data1[, 1] ))
#test_data1[, 1] <- as.numeric(as.character( test_data1[, 1]))
#colnames(test_data1)[1] <- 'deal'
#colnames(train_data1)[1] <- 'deal'
library(data.table)
library(scales)
tmp_DT = data.table(test_data1)
rank <- tmp_DT[, list(
  cnt = length(deal),
  cnt_resp = sum(deal),
  cnt_non_resp = sum(deal == 0)) ,
  by=deciles][order(-deciles)]
rank$rrate <- round (rank$cnt_resp / rank$cnt,4);
rank$cum_resp <- cumsum(rank$cnt_resp)
rank$cum_non_resp <- cumsum(rank$cnt_non_resp)
rank$cum_rel_resp <- round(rank$cum_resp / sum(rank$cnt_resp),4);
rank$cum_rel_non_resp <- round(rank$cum_non_resp /sum(rank$cnt_non_resp),4);
rank$ks <- abs(rank$cum_rel_resp - rank$cum_rel_non_resp);

library(scales)
rank$rrate <- percent(rank$rrate)
rank$cum_rel_resp <- percent(rank$cum_rel_resp)
#rank$cum_rel_non_resp <- percent(rank$cum_rel_non_resp)

View(rank)
```

```
#test_data[, 1] <- as.numeric(as.character( test_data[, 1]))
#colnames(test_data)[1] <- 'deal'

library(ROCR)
library(ineq)
pred_RF <- prediction(test_data1$predict.score[,2], test_data1$deal)
perf_new_RF<-ROCR::performance(pred_RF,"tpr","fpr")
plot(perf_new_RF,col="red", main="parameters_ROC")
abline(0,1, lty = 8, col = "grey")
```




```
KS <- max(attr(perf_new_RF, 'y.values')[[1]]-attr(perf_new_RF, 'x.values')[[1]])
KS
```

```
## [1] 0.8385388
```

```
## Area Under Curve
auc <-ROCR:: performance(pred_RF,"auc");
auc <- as.numeric(auc@y.values)
auc
```

```
## [1] 0.9708676
```

```
## Gini Coefficient
library(ineq)
gini = ineq(test_data1$predict.score[,2], type="Gini")
gini
```

```
## [1] 0.2341306
```

```
## Classification Error
with(test_data1,table(deal, predict.class))
```

```
##     predict.class
## deal  1  2
##    1 73  0
##    2 26 49
```

# Logstic regiress

```
#Logstic regiress
str(data_shark1)
```

```
## 'data.frame':    495 obs. of  523 variables:
##  $ deal            : num  0 1 1 0 0 1 0 0 0 1 ...
##  $ ratio           : int  14 10 15 25 10 15 20 20 10 10 ...
##  $ devic           : num  1 0 0 0 0 0 0 0 1 0 ...
##  $ retail          : num  0 2 0 0 0 1 0 0 0 0 ...
##  $ two             : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ new             : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ parent          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ young           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ children        : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ make            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ easi            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ littl           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ one             : num  0 0 1 0 0 1 0 0 0 0 ...
##  $ turn            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ experi          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ play            : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ provid          : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ posit           : num  0 0 1 0 0 0 0 0 0 0 ...
##  $ organ           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ pack            : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ move            : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ servic          : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ deliv           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ colleg          : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ women           : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ center          : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ room            : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ offer           : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ access          : num  0 0 0 0 1 0 0 0 0 0 ...
##  $ educ            : num  0 0 0 0 1 0 0 0 0 1 ...
##  $ first           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ pitch           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ shark           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ tank            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ perfect         : num  0 0 0 0 0 2 0 0 0 0 ...
##  $ line            : num  0 0 0 0 0 1 1 0 0 0 ...
##  $ focus           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ gourmet         : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ food            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ product         : num  0 0 0 0 0 2 0 0 0 0 ...
##  $ sold            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ across          : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ store           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ includ          : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ dress           : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ mani            : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ design          : num  0 0 0 0 0 1 0 0 0 1 ...
```

```
##  $ flavor          : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ shark_tank      : num  0 0 0 0 0 1 0 0 0 0 ...
##  $ mix             : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ cloth           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ look            : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ becom           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ big             : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ brand           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ activ           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ sport           : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ apparel         : num  0 0 0 0 0 0 1 0 0 0 ...
##  $ attach          : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ hold            : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ side            : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ screen          : num  0 0 0 0 0 0 0 1 0 0 ...
##  $ safeti          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ prevent         : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ start           : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ vehicl          : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ hous            : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ get             : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ student         : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ learn           : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ classic         : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ work            : num  0 0 0 0 0 0 0 0 0 1 ...
##  $ cook            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ household       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ item            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ made            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ recycl          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ book            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ help            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ find            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ fit             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ fun             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ minut           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ effici          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ entertain       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ us              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ set             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ card            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ combin          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ keep            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ kid             : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ long            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ school          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ develop         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ model           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ compani         : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ profession      : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ remov           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ much            : num  0 0 0 0 0 0 0 0 0 0 ...
##   [list output truncated]
```

```
summary (data_shark1)
```

```
##       deal            ratio            devic             retail
##  Min.   :0.0000   Min.   :  2.00   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.: 10.00   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :1.0000   Median : 15.00   Median :0.00000   Median :0.00000
##  Mean   :0.5071   Mean   : 17.44   Mean   :0.05051   Mean   :0.03434
##  3rd Qu.:1.0000   3rd Qu.: 20.00   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :100.00   Max.   :2.00000   Max.   :2.00000
##       two               new              parent           young
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.03636   Mean   :0.04646   Mean   :0.0202   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :2.00000   Max.   :2.0000   Max.   :2.00000
##     children            make             easi             littl
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.05455   Mean   :0.2162   Mean   :0.05051   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :3.0000   Max.   :2.00000   Max.   :2.00000
##       one               turn            experi             play
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.07677   Mean   :0.0303   Mean   :0.01414   Mean   :0.0404
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :4.00000   Max.   :2.0000   Max.   :1.00000   Max.   :6.0000
##      provid            posit             organ              pack
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.05051   Mean   :0.01212   Mean   :0.03636   Mean   :0.02222
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :1.00000   Max.   :1.00000
##       move             servic            deliv            colleg
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.01616   Mean   :0.06061   Mean   :0.0202   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :2.00000   Max.   :1.0000   Max.   :2.00000
##      women             center             room             offer
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02828   Mean   :0.01616   Mean   :0.01616   Mean   :0.05859
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
```

```
##  Max.   :2.00000   Max.   :2.00000   Max.   :1.00000   Max.   :2.00000
##      access            educ            first            pitch
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.01616   Mean   :0.02424   Mean   :0.01212
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :2.00000   Max.   :1.00000
##      shark             tank            perfect           line
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.01616   Mean   :0.01616   Mean   :0.0202   Mean   :0.06465
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :1.00000   Max.   :2.0000   Max.   :2.00000
##      focus           gourmet            food            product
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.01818   Mean   :0.01818   Mean   :0.04646   Mean   :0.1313
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :4.00000   Max.   :3.00000   Max.   :3.0000
##       sold            across            store            includ
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02424   Mean   :0.01414   Mean   :0.04646   Mean   :0.05253
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :3.00000   Max.   :2.00000
##      dress             mani            design           flavor
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.0000   Median :0.00000
##  Mean   :0.01616   Mean   :0.0202   Mean   :0.1475   Mean   :0.06263
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :3.00000   Max.   :1.0000   Max.   :3.0000   Max.   :3.00000
##    shark_tank           mix             cloth             look
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.01212   Mean   :0.0404   Mean   :0.04848   Mean   :0.05657
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :6.0000   Max.   :3.00000   Max.   :2.00000
##      becom             big             brand             activ
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01414   Mean   :0.01414   Mean   :0.03838   Mean   :0.02626
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
```

```
##  Max.   :1.00000   Max.   :1.00000   Max.   :2.00000   Max.   :2.00000
##       sport             apparel            attach              hold
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01818   Mean   :0.02424   Mean   :0.03232   Mean   :0.02424
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :3.00000
##       side              screen            safeti            prevent
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.01212   Mean   :0.01818   Mean   :0.01414   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :3.00000   Max.   :1.00000   Max.   :1.0000
##      start             vehicl             hous               get
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02626   Mean   :0.02222   Mean   :0.01616   Mean   :0.05051
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :4.00000   Max.   :2.00000   Max.   :2.00000
##     student            learn             classic            work
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.01818   Mean   :0.02626   Mean   :0.0101   Mean   :0.04848
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :3.00000   Max.   :3.00000   Max.   :1.0000   Max.   :2.00000
##       cook            household             item              made
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.0000
##  Mean   :0.01616   Mean   :0.01616   Mean   :0.0202   Mean   :0.1636
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :4.00000   Max.   :1.00000   Max.   :2.0000   Max.   :2.0000
##      recycl              book              help              find
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.01818   Mean   :0.01212   Mean   :0.08283   Mean   :0.0202
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :2.00000   Max.   :2.00000   Max.   :2.00000   Max.   :2.0000
##       fit               fun              minut             effici
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.05859   Mean   :0.05051   Mean   :0.02222   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
```

```
##  Max.   :4.00000   Max.   :2.00000   Max.   :3.00000   Max.   :1.0000
##    entertain            us                 set               card
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.01212   Mean   :0.01818   Mean   :0.03232
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :1.00000   Max.   :1.00000   Max.   :6.00000
##      combin             keep              kid               long
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02626   Mean   :0.05455   Mean   :0.05859   Mean   :0.01616
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :2.00000   Max.   :1.00000
##      school           develop            model            compani
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.01414   Mean   :0.01818   Mean   :0.01616   Mean   :0.1212
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :2.00000   Max.   :1.00000   Max.   :1.00000   Max.   :2.0000
##    profession          remov              much              like
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.03434   Mean   :0.01818   Mean   :0.08889
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :2.00000   Max.   :3.00000
##       yet               babi              spend             time
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.02424   Mean   :0.03636   Mean   :0.0101   Mean   :0.05859
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :3.00000   Max.   :1.0000   Max.   :2.00000
##       can              rescu              size              quick
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.2141   Mean   :0.01212   Mean   :0.04444   Mean   :0.03434
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :4.0000   Max.   :1.00000   Max.   :2.00000   Max.   :2.00000
##      solut              add               anoth             layer
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0303   Mean   :0.02222   Mean   :0.01212   Mean   :0.01414
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
```

```
##  Max.   :1.0000   Max.   :2.00000   Max.   :1.00000   Max.   :1.00000
##     protect            surfac           machin           maintain
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.04848   Mean   :0.0101   Mean   :0.01212   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :3.00000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
##      throw             wash             easier             tri
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.01414   Mean   :0.02828   Mean   :0.01212
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :2.00000   Max.   :1.00000   Max.   :1.00000
##      direct            bright           stylish             good
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01212   Mean   :0.01818   Mean   :0.01818   Mean   :0.01818
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :5.00000   Max.   :2.00000   Max.   :1.00000
##       own              coffe              high            caffein
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.03232   Mean   :0.05455   Mean   :0.01212
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :7.00000   Max.   :1.00000   Max.   :1.00000
##       use               toy              base               buy
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.1576   Mean   :0.0303   Mean   :0.03232   Mean   :0.02828
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :4.0000   Max.   :3.0000   Max.   :1.00000   Max.   :2.00000
##       back              sell              gift              year
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0303   Mean   :0.04242   Mean   :0.02828   Mean   :0.02828
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :2.00000   Max.   :3.00000   Max.   :1.00000
##      inspir           accessori           color             onlin
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.04242   Mean   :0.03434   Mean   :0.06061
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
```

```
##  Max.   :1.00000   Max.   :2.00000   Max.   :2.00000   Max.   :2.00000
##       user            toward            well              bar
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.04848   Mean   :0.01212   Mean   :0.03232   Mean   :0.04242
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :3.00000   Max.   :1.00000   Max.   :1.00000   Max.   :5.00000
##       safe            enjoy           unlik             market
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.05859   Mean   :0.0303   Mean   :0.01818   Mean   :0.02828
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :2.0000   Max.   :1.00000   Max.   :1.00000
##       low              less             sugar             damag
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.01616   Mean   :0.02222   Mean   :0.01212
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :4.00000   Max.   :2.00000   Max.   :3.00000   Max.   :1.00000
##     ingredi            natur             test             ultim
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.03636   Mean   :0.08081   Mean   :0.01212   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :4.00000   Max.   :4.00000   Max.   :1.00000   Max.   :1.0000
##       just              band             weight           handl
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.0000
##  Mean   :0.09091   Mean   :0.01414   Mean   :0.0202   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :2.00000   Max.   :2.00000   Max.   :2.0000   Max.   :1.0000
##      famili           fashion           engin             bodi
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.0000   Median :0.00000
##  Mean   :0.03232   Mean   :0.0404   Mean   :0.0101   Mean   :0.02828
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :1.0000   Max.   :1.0000   Max.   :3.00000
##      order             energi            whole             custom
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02424   Mean   :0.01616   Mean   :0.01212   Mean   :0.06061
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
```

```
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :4.00000
##      choos               go             person          made_order
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.01616   Mean   :0.04646   Mean   :0.02828   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :2.00000   Max.   :1.00000   Max.   :1.0000
##  natur_ingredi         beer             bottl             award
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.01212   Mean   :0.04849   Mean   :0.01212
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :7.00000   Max.   :1.00000
##       win              materi            around             leg
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.01212   Mean   :0.02828   Mean   :0.02828   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :2.00000   Max.   :2.00000   Max.   :2.00000   Max.   :1.0000
##     program             danc              live            attract
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.01414   Mean   :0.0101   Mean   :0.02626   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.0000   Max.   :2.00000   Max.   :1.00000
##       five              featur            stick              free
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02424   Mean   :0.04646   Mean   :0.01616   Mean   :0.06667
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :3.00000   Max.   :2.00000   Max.   :3.00000   Max.   :4.00000
##       pair              three           instead             uniqu
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.0303   Mean   :0.02424   Mean   :0.03636
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :2.0000   Max.   :1.00000   Max.   :2.00000
##       belt              allow              stuf              anim
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.02626   Mean   :0.06061   Mean   :0.0101   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
```

```
##  Max.   :5.00000   Max.   :2.00000   Max.   :1.0000   Max.   :1.00000
##    look_like          special        special_design       beverag
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.03434   Mean   :0.01212   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##       dog               golf              need             cover
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.03434   Mean   :0.01414   Mean   :0.04646   Mean   :0.03434
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :3.00000   Max.   :1.00000   Max.   :2.00000   Max.   :4.00000
##       run              replac             full             chain
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02222   Mean   :0.01414   Mean   :0.03232   Mean   :0.01212
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :2.00000   Max.   :2.00000   Max.   :1.00000
##      blend             season           portabl             ball
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02626   Mean   :0.01212   Mean   :0.01616   Mean   :0.01616
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :1.00000   Max.   :1.00000   Max.   :4.00000
##      patent            place            motorcycl           kind
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.03232   Mean   :0.02828   Mean   :0.0101   Mean   :0.01818
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :2.00000   Max.   :1.0000   Max.   :1.00000
##       eco              friend          eco_friend         individu
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02424   Mean   :0.05253   Mean   :0.01616   Mean   :0.01212
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :4.00000   Max.   :4.00000   Max.   :1.00000   Max.   :1.00000
##      packag            clean            easili            bicycl
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.02828   Mean   :0.0303   Mean   :0.02828   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
```

```
##  Max.   :3.00000   Max.   :4.0000   Max.   :2.00000   Max.   :1.0000
##       face             storag            pocket             shop
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.01616   Mean   :0.01616   Mean   :0.0202   Mean   :0.01616
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :2.00000   Max.   :4.0000   Max.   :1.00000
##       way              save             mobil              ice
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.06667   Mean   :0.01818   Mean   :0.02626   Mean   :0.0202
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :2.00000   Max.   :1.00000   Max.   :1.00000   Max.   :3.0000
##      cream             busi           ice_cream            ever
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0202   Mean   :0.04242   Mean   :0.01212   Mean   :0.01414
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.0000   Max.   :2.00000   Max.   :2.00000   Max.   :1.00000
##     whether           regular             stand             print
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01212   Mean   :0.01818   Mean   :0.01616   Mean   :0.01818
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :1.00000   Max.   :2.00000   Max.   :3.00000
##       see              sinc              take              still
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.01616   Mean   :0.04444   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :2.00000   Max.   :1.00000
##       even              show            manufactur           cost
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.08889   Mean   :0.01616   Mean   :0.01212   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :2.00000   Max.   :2.00000   Max.   :1.00000   Max.   :1.0000
##       age              shape             creat             encourag
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02424   Mean   :0.02828   Mean   :0.05051   Mean   :0.01818
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
```

```
##  Max.   :2.00000   Max.   :2.00000   Max.   :2.00000   Max.   :1.00000
##       pre               plan              usa             made_usa
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01212   Mean   :0.01414   Mean   :0.02424   Mean   :0.02222
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :2.00000   Max.   :1.00000   Max.   :1.00000
##       hand              want              thing             practic
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.0404   Mean   :0.03636   Mean   :0.01818   Mean   :0.0101
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :5.0000   Max.   :2.00000   Max.   :1.00000   Max.   :1.0000
##      receiv            give              home              shoe
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.03636   Mean   :0.05455   Mean   :0.02828
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :2.00000   Max.   :2.00000
##       sale              parti             stay              without
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.0202   Mean   :0.01818   Mean   :0.06465
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.0000   Max.   :5.00000   Max.   :2.00000
##     struggl           plastic             men              signatur
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.0101   Mean   :0.0202   Mean   :0.02626   Mean   :0.0101
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :4.00000   Max.   :1.0000
##      power             secur             strong            magnet
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.04242   Mean   :0.01818   Mean   :0.01212   Mean   :0.01616
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :1.00000   Max.   :1.00000   Max.   :2.00000
##      shirt             produc            expand             suit
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01818   Mean   :0.03636   Mean   :0.01414   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
```

```
##  Max.   :4.00000   Max.   :5.00000   Max.   :1.00000   Max.   :1.00000
##      money             clip              tie              travel
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02828   Mean   :0.01414   Mean   :0.02424   Mean   :0.02424
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :4.00000   Max.   :2.00000   Max.   :6.00000   Max.   :1.00000
##       bag             premium             wine             glass
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02424   Mean   :0.03838   Mean   :0.02626   Mean   :0.01818
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :2.00000   Max.   :3.00000   Max.   :2.00000
##       cup             flagship           origin        flagship_product
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0202   Mean   :0.01616   Mean   :0.01212   Mean   :0.01616
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :3.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##    subscript           fresh             bring              self
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.02222   Mean   :0.02626   Mean   :0.01616
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##       air               come              reduc             effect
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.02222   Mean   :0.06061   Mean   :0.0202   Mean   :0.02828
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :3.00000   Max.   :2.00000   Max.   :1.0000   Max.   :1.00000
##      super          lightweight          ultra              durabl
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01818   Mean   :0.01212   Mean   :0.01212   Mean   :0.01616
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##       ride              tradit            simpli             bike
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :0.0000   Median :0.0000   Median :0.0000
##  Mean   :0.02222   Mean   :0.0202   Mean   :0.0101   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
```

```
##  Max.   :3.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
##       feet             fulli             readi             resist
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01212   Mean   :0.01414   Mean   :0.01616   Mean   :0.01616
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :1.00000   Max.   :2.00000
##       mom              qualiti             care              craft
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.01414   Mean   :0.03636   Mean   :0.01414   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :3.00000   Max.   :1.00000   Max.   :3.00000   Max.   :1.0000
##       piec               serv              found             current
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01212   Mean   :0.03232   Mean   :0.01414   Mean   :0.01212
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :1.00000   Max.   :1.00000
##      expens             sourc             sustain              wood
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.01212   Mean   :0.01212   Mean   :0.01414   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :2.00000   Max.   :1.00000   Max.   :1.00000   Max.   :1.0000
##   high_qualiti         avail              choic               let
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.0000
##  Mean   :0.0101   Mean   :0.03434   Mean   :0.01212   Mean   :0.0101
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.0000
##     kid_can            anyon               pet               smell
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.02222   Mean   :0.02222   Mean   :0.02424
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :2.00000   Max.   :3.00000   Max.   :5.00000
##      differ             spray             chemic            continu
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.01818   Mean   :0.0303   Mean   :0.01414   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
```

```
##  Max.   :1.00000   Max.   :8.0000   Max.   :1.00000   Max.   :1.0000
##     can_buy           countri           style             truck
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.01616   Mean   :0.03232   Mean   :0.02222
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :3.00000   Max.   :2.00000   Max.   :4.00000
##       stop              seat             addit            preserv
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.01818   Mean   :0.01212   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :4.00000   Max.   :2.00000   Max.   :2.00000
##       lip              light              car              window
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01818   Mean   :0.03838   Mean   :0.01616   Mean   :0.01212
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :3.00000   Max.   :4.00000   Max.   :3.00000   Max.   :2.00000
##      small              case              gear              skin
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.02828   Mean   :0.01616   Mean   :0.0101   Mean   :0.03232
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.0000   Max.   :3.00000
##      appli             built              cake              recip
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01212   Mean   :0.04242   Mean   :0.02424   Mean   :0.01818
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :7.00000   Max.   :3.00000
##       ship            system             pad              train
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.06061   Mean   :0.02626   Mean   :0.02626
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :2.00000   Max.   :2.00000   Max.   :4.00000
##      video              game              altern           cours
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.01616   Mean   :0.02222   Mean   :0.0202   Mean   :0.01818
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
```

```
##  Max.   :1.00000   Max.   :2.00000   Max.   :1.0000   Max.   :1.00000
##      among              non               wed              event
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01212   Mean   :0.02222   Mean   :0.01212   Mean   :0.01616
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :1.00000   Max.   :2.00000
##       fast              fire             connect            tool
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.01212   Mean   :0.02424   Mean   :0.0101   Mean   :0.02828
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :5.00000   Max.   :1.0000   Max.   :3.00000
##       kit               box             success           healthi
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.02626   Mean   :0.0303   Mean   :0.01414   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :3.00000   Max.   :3.0000   Max.   :2.00000   Max.   :1.00000
##      month            display          interchang        allow_user
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01414   Mean   :0.01212   Mean   :0.01616   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :1.00000   Max.   :1.00000
##      theme          make_fashion        water             contain
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.01414   Mean   :0.0101   Mean   :0.07475   Mean   :0.02626
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :1.0000   Max.   :6.00000   Max.   :1.00000
##       top               eas              websit            peopl
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.01818   Mean   :0.0101   Mean   :0.01212   Mean   :0.0404
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
##     purchas           partner             larg           prospect
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.0000
##  Mean   :0.01212   Mean   :0.01212   Mean   :0.0101   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000
```

```
##  Max.   :2.00000   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
##    transform          carri              bed              magic
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0202   Mean   :0.01212   Mean   :0.02222   Mean   :0.01414
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :3.00000   Max.   :2.00000
##       day              electr            right            music
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.0000
##  Mean   :0.02424   Mean   :0.01818   Mean   :0.0101   Mean   :0.0202
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :2.00000   Max.   :2.00000   Max.   :1.0000   Max.   :3.0000
##       real             favorit          instruct           open
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.0000   Median :0.00000
##  Mean   :0.02626   Mean   :0.0101   Mean   :0.0101   Mean   :0.02222
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :2.00000   Max.   :1.0000   Max.   :1.0000   Max.   :2.00000
##      point            serious           butter             oil
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.01414   Mean   :0.0101   Mean   :0.03434   Mean   :0.0202
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :1.0000   Max.   :6.00000   Max.   :3.0000
##      gluten             now               name            insid
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.01212   Mean   :0.02222   Mean   :0.0202   Mean   :0.01212
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.0000   Max.   :2.00000
##      inflat           chang            better           comfort
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.0202   Mean   :0.02222   Mean   :0.02222
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :2.0000   Max.   :2.00000   Max.   :2.00000
##      adjust           varieti            best             space
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.01212   Mean   :0.0202   Mean   :0.02222   Mean   :0.01818
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
```

```
##  Max.   :1.00000   Max.   :1.0000   Max.   :2.00000   Max.   :1.00000
##       flat           creativ            chocol            corpor
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.01818   Mean   :0.02222   Mean   :0.01212
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :4.00000   Max.   :1.00000
##      improv           equip          technolog          although
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.0101   Mean   :0.0101   Mean   :0.01212   Mean   :0.0101
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
##      number            might            conveni            rest
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.01414   Mean   :0.01818   Mean   :0.0101   Mean   :0.01616
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :1.0000   Max.   :1.00000
##      alway             rather           smart            kitchen
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.01212   Mean   :0.0101   Mean   :0.01818   Mean   :0.01212
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :1.0000   Max.   :3.00000   Max.   :1.00000
##      treat              cut              paper             accord
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02626   Mean   :0.01818   Mean   :0.01616   Mean   :0.01212
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :3.00000   Max.   :2.00000   Max.   :3.00000   Max.   :1.00000
##     can_use           price             togeth            everi
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.01818   Mean   :0.01414   Mean   :0.02828
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :1.00000
##       rang         premium_qualiti        meal            portion
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.02222   Mean   :0.01212   Mean   :0.01414   Mean   :0.01616
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
```

```
##  Max.   :1.00000   Max.   :1.00000   Max.   :3.00000   Max.   :1.00000
##       wast           workout            secret            visibl
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0202   Mean   :0.01414   Mean   :0.01212   Mean   :0.01212
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :3.0000   Max.   :2.00000   Max.   :2.00000   Max.   :1.00000
##      wearer            mother            drink             put
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.0000
##  Mean   :0.01414   Mean   :0.01212   Mean   :0.0303   Mean   :0.0202
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :2.00000   Max.   :2.0000   Max.   :2.0000
##      thank             wheel             inch              nut
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.02828   Mean   :0.01414   Mean   :0.03232
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :6.00000   Max.   :2.00000   Max.   :5.00000
##      singl             peanut            social             call
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.01414   Mean   :0.01212   Mean   :0.01212   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :1.00000   Max.   :1.00000
##    smartphon           build             app              reusabl
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.02424   Mean   :0.0202   Mean   :0.03434   Mean   :0.01616
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :3.00000   Max.   :3.0000   Max.   :3.00000   Max.   :3.00000
##      sturdi            phone             grow             afford
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.0101   Mean   :0.02424   Mean   :0.0101   Mean   :0.02222
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :5.00000   Max.   :1.0000   Max.   :1.00000
##     content            consum            pay               may
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.0000   Median :0.00000
##  Mean   :0.01616   Mean   :0.0202   Mean   :0.0101   Mean   :0.01212
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000
```

```
##  Max.   :2.00000   Max.   :2.0000   Max.   :1.0000   Max.   :1.00000
##       feel             form              ad               roll
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.01616   Mean   :0.02424   Mean   :0.01818
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000   Max.   :3.00000
##    california         balanc           control           support
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.01616   Mean   :0.01818   Mean   :0.01212
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :2.00000   Max.   :2.00000   Max.   :1.00000
##       cool             heat           function           remot
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.01818   Mean   :0.02222   Mean   :0.01414
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :4.00000   Max.   :1.00000   Max.   :1.00000
##       tast             hair           outdoor            innov
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.02626   Mean   :0.01616   Mean   :0.01414
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :6.00000   Max.   :2.00000   Max.   :1.00000
##      invent            love             sound            snack
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.01212   Mean   :0.01414   Mean   :0.0202   Mean   :0.01212
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :3.0000   Max.   :1.00000
##      someth            result           process           delici
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.01212   Mean   :0.01616   Mean   :0.0101   Mean   :0.01414
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
##  Max.   :1.00000   Max.   :2.00000   Max.   :1.0000   Max.   :1.00000
##     versatil           goe              simpl           complet
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.0000   Median :0.00000
##  Mean   :0.0101   Mean   :0.01212   Mean   :0.0303   Mean   :0.01212
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000
```

```
##  Max.   :1.0000   Max.   :2.00000   Max.   :2.0000   Max.   :1.00000
##     mission            detail            old              great
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.0101   Mean   :0.01212   Mean   :0.02424
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :3.00000   Max.   :1.0000   Max.   :1.00000   Max.   :1.00000
##      audio            promot           iphon             nation
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.0101   Mean   :0.01414   Mean   :0.01212
##  3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.0000   Max.   :3.00000   Max.   :2.00000
##       team              duo             behind            tablet
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.00000   Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.01616   Mean   :0.0101   Mean   :0.01212   Mean   :0.01616
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :4.00000   Max.   :1.0000   Max.   :1.00000   Max.   :4.00000
##       meet             select            world           respons
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :0.0000
##  Mean   :0.01212   Mean   :0.01616   Mean   :0.0202   Mean   :0.0101
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :1.00000   Max.   :2.00000   Max.   :2.0000   Max.   :1.0000
##      beach            typic           companion
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.00000
##  1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.00000
##  Median :0.0000   Median :0.00000   Median :0.00000
##  Mean   :0.0101   Mean   :0.01616   Mean   :0.01212
##  3rd Qu.:0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.00000
```

```
# partition the data
set.seed(1000)
library(dplyr)
library(caTools)
spl = sample.split(data_shark1$deal, SplitRatio=0.70)
train1 = subset(data_shark1, spl ==T)
test1 = subset(data_shark1, spl==F)
dim(train1)
```

```
## [1] 347 523
```

```
# create a model
LRmodel_new = glm(deal ~ ., data = train1, family = binomial)
```

```
## Warning: glm.fit: algorithm did not converge
```

```
summary(LRmodel_new)
```

```
##
## Call:
## glm(formula = deal ~ ., family = binomial, data = train1)
##
## Deviance Residuals:
##   [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [24]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [47]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [70]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##  [93]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [116]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [139]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [162]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [185]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [208]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [231]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [254]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [277]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [300]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [323]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## [346]  0  0
##
## Coefficients: (176 not defined because of singularities)
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.328e+02  1.284e+06       0        1
## ratio             1.063e+01  1.007e+05       0        1
## devic            -4.251e+01  4.616e+05       0        1
## retail            2.152e+04  1.282e+08       0        1
## two              -4.885e+04  2.916e+08       0        1
## new               5.867e+03  3.541e+07       0        1
## parent           -6.679e+03  4.359e+07       0        1
## young            -9.693e+04  5.750e+08       0        1
## children         -1.786e+02  1.969e+06       0        1
## make              5.873e+03  3.544e+07       0        1
## easi             -2.694e+03  1.805e+07       0        1
## littl            -4.433e+04  2.788e+08       0        1
## one              -7.514e+03  4.601e+07       0        1
## turn             -5.313e+01  7.122e+05       0        1
## experi           -5.346e+04  3.185e+08       0        1
## play             -3.389e+03  2.032e+07       0        1
## provid           -3.553e+03  2.127e+07       0        1
## posit             2.626e+04  1.570e+08       0        1
## organ            -6.376e+01  1.037e+06       0        1
## pack             -3.893e+04  2.291e+08       0        1
## move             -7.346e+04  4.428e+08       0        1
## servic            5.980e+03  3.521e+07       0        1
## deliv             1.892e+04  1.105e+08       0        1
## colleg            3.747e+04  2.205e+08       0        1
```

```
## women            9.271e+03  5.534e+07       0        1
## center          -1.085e+04  6.673e+07       0        1
## room            -1.599e+03  1.788e+07       0        1
## offer           -4.869e+03  3.205e+07       0        1
## access          -1.041e+03  7.307e+06       0        1
## educ             1.836e+04  1.129e+08       0        1
## first            3.248e-10  5.036e+05       0        1
## pitch            1.274e+04  7.623e+07       0        1
## shark           -9.380e+03  6.292e+07       0        1
## tank            -2.125e+02  1.816e+06       0        1
## perfect         -1.741e+04  1.053e+08       0        1
## line             2.801e+03  1.706e+07       0        1
## focus           -8.343e+03  5.035e+07       0        1
## gourmet          7.479e+03  4.525e+07       0        1
## food             7.728e+03  4.559e+07       0        1
## product         -4.400e+03  2.582e+07       0        1
## sold            -9.207e+03  5.557e+07       0        1
## across          -5.812e+04  3.479e+08       0        1
## store           -1.283e+02  4.177e+06       0        1
## includ           4.710e+03  3.088e+07       0        1
## dress            1.685e+04  1.007e+08       0        1
## mani            -7.527e+03  5.318e+07       0        1
## design          -2.742e+03  1.623e+07       0        1
## flavor          -1.063e+02  8.723e+05       0        1
## shark_tank      -1.235e+04  7.491e+07       0        1
## mix              2.848e+03  1.671e+07       0        1
## cloth            1.116e+04  6.687e+07       0        1
## look            -6.032e+03  3.642e+07       0        1
## becom            4.760e+03  2.762e+07       0        1
## big             -2.061e+04  1.227e+08       0        1
## brand           -5.729e+03  3.377e+07       0        1
## activ            4.560e+03  2.714e+07       0        1
## sport            2.537e+03  1.628e+07       0        1
## apparel          3.601e+03  2.144e+07       0        1
## attach           4.400e+03  2.588e+07       0        1
## hold             4.814e+03  2.778e+07       0        1
## side             2.134e+03  1.266e+07       0        1
## screen          -1.145e+04  6.679e+07       0        1
## safeti          -2.194e+04  1.313e+08       0        1
## prevent          1.301e+04  7.584e+07       0        1
## start            9.030e+03  5.914e+07       0        1
## vehicl          -5.313e+01  8.723e+05       0        1
## hous            -2.693e+04  1.628e+08       0        1
## get             -4.509e+02  6.203e+06       0        1
## student          2.855e+04  1.703e+08       0        1
## learn            3.551e+04  2.122e+08       0        1
## classic         -7.379e+04  4.416e+08       0        1
## work            -4.398e+03  2.702e+07       0        1
## cook            -1.700e+02  2.174e+06       0        1
```

```
## household       -1.242e+04  7.444e+07       0        1
## item            -7.953e+03  4.995e+07       0        1
## made             6.376e+01  1.061e+06       0        1
## recycl           5.313e+01  5.036e+05       0        1
## book             1.060e+04  6.377e+07       0        1
## help            -3.068e+03  1.843e+07       0        1
## find            -2.137e+04  1.258e+08       0        1
## fit              2.795e+03  1.671e+07       0        1
## fun              7.745e+03  4.537e+07       0        1
## minut           -3.174e+03  2.219e+07       0        1
## effici           6.404e+01  7.898e+06       0        1
## entertain       -4.650e+03  2.881e+07       0        1
## us               3.254e+04  1.939e+08       0        1
## set              2.661e+03  1.690e+07       0        1
## card             1.402e+04  8.511e+07       0        1
## combin          -1.919e+04  1.147e+08       0        1
## keep             6.272e+03  3.731e+07       0        1
## kid              2.679e+03  1.623e+07       0        1
## long            -1.086e+05  6.488e+08       0        1
## school          -2.673e+04  1.583e+08       0        1
## develop          2.195e+04  1.311e+08       0        1
## model            8.308e+03  4.983e+07       0        1
## compani         -5.980e+03  3.519e+07       0        1
## profession       3.479e-10  5.036e+05       0        1
## remov            2.488e+03  1.728e+07       0        1
## much            -1.570e+04  9.304e+07       0        1
## like            -7.147e+03  4.364e+07       0        1
## yet              4.600e+04  2.732e+08       0        1
## babi            -2.765e+03  1.700e+07       0        1
## spend           -6.629e+03  3.894e+07       0        1
## time             3.953e+03  2.415e+07       0        1
## can              4.465e+03  2.907e+07       0        1
## rescu           -2.339e+04  1.427e+08       0        1
## size            -3.447e+03  2.043e+07       0        1
## quick           -5.450e+03  3.637e+07       0        1
## solut            2.421e+03  1.394e+07       0        1
## add             -1.493e+02  2.237e+07       0        1
## anoth            6.958e+04  4.153e+08       0        1
## layer           -6.998e+03  4.929e+07       0        1
## protect         -1.275e+02  1.403e+06       0        1
## surfac          -2.658e+03  1.882e+07       0        1
## machin           1.960e+03  1.412e+07       0        1
## maintain         2.934e+04  1.773e+08       0        1
## throw            4.567e+04  2.708e+08       0        1
## wash            -1.132e+04  6.956e+07       0        1
## easier          -1.673e+04  1.012e+08       0        1
## tri             -3.813e+04  2.261e+08       0        1
## direct           2.994e+04  1.806e+08       0        1
## bright          -4.426e+04  2.656e+08       0        1
```

```
## stylish          -1.315e+04  7.748e+07       0        1
## good              3.553e+02  8.520e+06       0        1
## own               8.126e+03  4.733e+07       0        1
## coffe             2.285e+03  1.341e+07       0        1
## high              6.122e+03  3.911e+07       0        1
## caffein          -2.771e+04  1.651e+08       0        1
## use              -2.390e+03  1.478e+07       0        1
## toy               1.923e+04  1.141e+08       0        1
## base             -5.484e+03  3.332e+07       0        1
## buy              -7.551e+04  4.452e+08       0        1
## back              8.919e-11  5.036e+05       0        1
## sell              6.104e+03  3.569e+07       0        1
## gift             -9.965e+03  5.876e+07       0        1
## year              2.415e+04  1.468e+08       0        1
## inspir           -2.742e+03  1.661e+07       0        1
## accessori         1.413e+04  8.535e+07       0        1
## color            -2.135e+04  1.258e+08       0        1
## onlin            -1.708e+03  1.414e+07       0        1
## user              2.125e+01  7.737e+05       0        1
## toward           -2.108e+04  1.298e+08       0        1
## well             -4.265e+03  2.559e+07       0        1
## bar              -2.496e+03  1.435e+07       0        1
## safe              5.349e+03  3.150e+07       0        1
## enjoy             2.785e+03  1.728e+07       0        1
## unlik            -6.031e+04  3.574e+08       0        1
## market           -1.784e+04  1.062e+08       0        1
## low               1.730e+04  1.024e+08       0        1
## less              3.747e+04  2.205e+08       0        1
## sugar            -5.517e+04  3.301e+08       0        1
## damag            -3.335e+04  1.932e+08       0        1
## ingredi           3.588e+03  2.079e+07       0        1
## natur            -1.567e+03  1.423e+07       0        1
## test              2.281e+04  1.357e+08       0        1
## ultim            -6.939e+03  4.337e+07       0        1
## just              2.625e+02  3.680e+06       0        1
## band             -1.063e+02  8.723e+05       0        1
## weight            6.155e+03  4.054e+07       0        1
## handl            -5.316e+04  3.173e+08       0        1
## famili            5.313e+01  5.036e+05       0        1
## fashion          -6.191e+03  3.781e+07       0        1
## engin             2.345e-10  5.036e+05       0        1
## bodi             -1.063e+02  8.723e+05       0        1
## order            -1.067e+04  7.116e+07       0        1
## energi            1.103e+04  6.594e+07       0        1
## whole            -1.201e+05  7.194e+08       0        1
## custom           -1.381e+02  1.439e+06       0        1
## choos            -1.162e+03  2.783e+07       0        1
## go               -1.107e+04  6.637e+07       0        1
## person           -1.273e+04  7.868e+07       0        1
```

```
## made_order        1.119e+03  4.504e+07       0        1
## natur_ingredi     1.324e+05  7.865e+08       0        1
## beer             -2.379e+03  1.408e+07       0        1
## bottl            -1.052e+04  6.313e+07       0        1
## award             5.530e+04  3.286e+08       0        1
## win              -5.695e+04  3.385e+08       0        1
## materi            5.224e+03  3.124e+07       0        1
## around            2.253e+03  1.328e+07       0        1
## leg              -4.958e+03  3.002e+07       0        1
## program           8.057e+03  4.968e+07       0        1
## danc             -1.149e+04  7.076e+07       0        1
## live              9.077e+03  5.538e+07       0        1
## attract           1.805e+04  1.098e+08       0        1
## five             -1.811e+04  1.092e+08       0        1
## featur           -3.771e+03  2.264e+07       0        1
## stick            -7.834e+03  4.653e+07       0        1
## free             -1.063e+01  8.487e+05       0        1
## pair              5.029e+04  2.999e+08       0        1
## three            -6.099e+02  5.051e+06       0        1
## instead          -1.113e+03  7.895e+06       0        1
## uniqu             3.984e+03  2.407e+07       0        1
## belt              5.180e+03  3.140e+07       0        1
## allow             3.431e+03  2.077e+07       0        1
## stuf              4.750e+03  3.901e+07       0        1
## anim             -2.992e+04  1.822e+08       0        1
## look_like         7.777e+04  4.617e+08       0        1
## special           7.706e+03  4.602e+07       0        1
## special_design   -1.444e+04  8.665e+07       0        1
## beverag          -7.228e+03  4.387e+07       0        1
## dog              -7.781e+03  4.585e+07       0        1
## golf             -4.621e+03  2.758e+07       0        1
## need              5.916e+03  3.506e+07       0        1
## cover             1.063e+01  7.193e+05       0        1
## run               2.988e+03  1.890e+07       0        1
## replac           -7.658e+03  4.731e+07       0        1
## full              4.516e+02  8.923e+06       0        1
## chain             6.376e+01  3.743e+06       0        1
## blend             8.362e+03  5.410e+07       0        1
## season           -1.993e+04  1.197e+08       0        1
## portabl          -4.368e+04  2.594e+08       0        1
## ball              4.798e+04  2.847e+08       0        1
## patent            1.934e+04  1.177e+08       0        1
## place            -7.652e+03  4.973e+07       0        1
## motorcycl        -7.066e+02  8.748e+06       0        1
## kind             -5.973e+03  3.643e+07       0        1
## eco              -7.239e+03  6.375e+07       0        1
## friend           -5.754e+02  3.917e+06       0        1
## eco_friend        9.836e+03  7.644e+07       0        1
## individu         -1.000e+04  6.066e+07       0        1
```

```
## packag          -2.363e+03  1.395e+07       0        1
## clean            4.255e+03  2.541e+07       0        1
## easili           8.098e+03  4.853e+07       0        1
## bicycl           1.656e+03  1.435e+07       0        1
## face             2.635e+03  1.522e+07       0        1
## storag           2.063e+03  1.244e+07       0        1
## pocket          -4.932e+03  2.907e+07       0        1
## shop            -1.063e+01  3.563e+06       0        1
## way             -1.987e+03  1.178e+07       0        1
## save             6.293e+04  3.724e+08       0        1
## mobil           -1.063e+01  5.136e+05       0        1
## ice              5.101e+04  3.044e+08       0        1
## cream           -3.916e+04  2.346e+08       0        1
## busi            -2.474e+03  1.458e+07       0        1
## ice_cream       -9.527e+03  6.141e+07       0        1
## ever            -2.315e+04  1.425e+08       0        1
## whether          8.272e+03  4.697e+07       0        1
## regular         -6.996e+04  4.133e+08       0        1
## stand           -7.275e+04  4.356e+08       0        1
## print            1.876e+03  1.649e+07       0        1
## see              3.522e+04  2.143e+08       0        1
## sinc             2.544e+04  1.500e+08       0        1
## take            -1.063e+02  8.723e+05       0        1
## still            9.888e+03  6.601e+07       0        1
## even             4.102e+04  2.433e+08       0        1
## show             1.102e+04  6.621e+07       0        1
## manufactur      -4.317e+04  2.598e+08       0        1
## cost             5.327e+04  3.258e+08       0        1
## age             -2.679e+03  1.648e+07       0        1
## shape            9.817e+03  6.341e+07       0        1
## creat           -1.705e+04  1.010e+08       0        1
## encourag        -2.004e+04  1.178e+08       0        1
## pre              6.419e+03  3.821e+07       0        1
## plan             1.248e+04  7.504e+07       0        1
## usa             -1.834e+04  1.089e+08       0        1
## made_usa                NA         NA      NA       NA
## hand             1.063e+01  1.514e+06       0        1
## want            -2.066e+04  1.229e+08       0        1
## thing           -2.730e+03  2.089e+07       0        1
## practic          2.738e+04  1.661e+08       0        1
## receiv          -9.021e+04  5.444e+08       0        1
## give            -1.224e+04  7.296e+07       0        1
## home             1.063e+02  1.126e+06       0        1
## shoe            -5.572e+03  3.338e+07       0        1
## sale            -1.874e+04  1.128e+08       0        1
## parti            2.380e+04  1.425e+08       0        1
## stay            -1.670e+04  1.022e+08       0        1
## without         -1.038e+04  6.340e+07       0        1
## struggl          3.190e+04  1.962e+08       0        1
```

```
## plastic          -1.410e+04  8.496e+07       0        1
## men               1.242e+04  7.454e+07       0        1
## signatur          1.022e+05  6.138e+08       0        1
## power            -1.712e+03  1.074e+07       0        1
## secur            -8.839e+03  5.316e+07       0        1
## strong           -3.791e+04  2.251e+08       0        1
## magnet            1.606e+03  9.673e+06       0        1
## shirt            -2.688e+03  1.587e+07       0        1
## produc            4.400e+03  2.591e+07       0        1
## expand           -7.499e+03  4.528e+07       0        1
## suit              6.584e+04  3.885e+08       0        1
## money             7.125e+03  4.261e+07       0        1
## clip             -2.674e+03  1.723e+07       0        1
## tie              -2.307e+04  1.381e+08       0        1
## travel           -5.873e+03  3.459e+07       0        1
## bag              -2.128e+03  1.251e+07       0        1
## premium           4.225e+04  2.541e+08       0        1
## wine              1.594e+02  1.126e+06       0        1
## glass            -4.269e+04  2.531e+08       0        1
## cup               3.364e+04  1.989e+08       0        1
## flagship          5.968e+03  3.625e+07       0        1
## origin            1.556e+04  9.496e+07       0        1
## flagship_product         NA         NA      NA       NA
## subscript        -2.515e+04  1.490e+08       0        1
## fresh             4.251e+01  1.860e+06       0        1
## bring             4.251e+01  1.196e+06       0        1
## self              1.931e+04  1.144e+08       0        1
## air               1.203e+04  7.364e+07       0        1
## come              4.040e+03  2.385e+07       0        1
## reduc            -2.615e+04  1.564e+08       0        1
## effect            3.177e+04  1.859e+08       0        1
## super            -4.925e+04  2.981e+08       0        1
## lightweight       1.716e+04  1.051e+08       0        1
## ultra             9.401e+04  5.604e+08       0        1
## durabl            5.761e+03  3.495e+07       0        1
## ride              8.028e+02  5.042e+06       0        1
## tradit           -1.108e+04  6.719e+07       0        1
## simpli            7.530e+04  4.553e+08       0        1
## bike              3.508e+04  2.075e+08       0        1
## feet              1.789e+04  1.078e+08       0        1
## fulli             2.537e+04  1.523e+08       0        1
## readi            -5.472e+04  3.297e+08       0        1
## resist           -4.507e+03  2.682e+07       0        1
## mom               2.635e+03  1.561e+07       0        1
## qualiti           2.818e+03  1.698e+07       0        1
## care             -1.228e+04  8.524e+07       0        1
## craft             3.014e-10  5.036e+05       0        1
## piec              4.949e+04  2.931e+08       0        1
## serv              3.603e+04  2.164e+08       0        1
```

```
## found             5.720e+03  3.878e+07       0        1
## current          -1.044e+05  6.240e+08       0        1
## expens            6.999e+04  4.365e+08       0        1
## sourc            -4.168e+04  2.455e+08       0        1
## sustain           5.486e+04  3.261e+08       0        1
## wood              3.189e+04  1.854e+08       0        1
## high_qualiti      2.483e+04  1.465e+08       0        1
## avail             5.873e+03  3.435e+07       0        1
## choic            -1.963e+04  1.181e+08       0        1
## let               2.473e+04  1.483e+08       0        1
## kid_can          -6.753e+04  3.977e+08       0        1
## anyon            -2.051e+04  1.237e+08       0        1
## pet               4.700e+03  2.825e+07       0        1
## smell            -1.411e+04  8.344e+07       0        1
## differ            9.247e+04  5.511e+08       0        1
## spray             1.553e+02  2.072e+06       0        1
## chemic            2.442e+04  1.587e+08       0        1
## continu          -1.854e+03  1.572e+07       0        1
## can_buy           5.829e+04  3.429e+08       0        1
## countri           8.359e+03  5.392e+07       0        1
## style            -9.583e+03  5.613e+07       0        1
## truck            -4.412e+03  2.836e+07       0        1
## stop              1.460e+04  8.578e+07       0        1
## seat              1.545e+03  9.521e+06       0        1
## addit            -3.337e+04  2.007e+08       0        1
## preserv           1.333e+03  9.201e+06       0        1
## lip              -1.018e+04  6.279e+07       0        1
## light             1.063e+02  8.723e+05       0        1
## car                      NA         NA      NA       NA
## window                   NA         NA      NA       NA
## small                    NA         NA      NA       NA
## case                     NA         NA      NA       NA
## gear                     NA         NA      NA       NA
## skin                     NA         NA      NA       NA
## appli                    NA         NA      NA       NA
## built                    NA         NA      NA       NA
## cake                     NA         NA      NA       NA
## recip                    NA         NA      NA       NA
## ship                     NA         NA      NA       NA
## system            1.047e-10  5.036e+05       0        1
## pad                      NA         NA      NA       NA
## train                    NA         NA      NA       NA
## video                    NA         NA      NA       NA
## game                     NA         NA      NA       NA
## altern                   NA         NA      NA       NA
## cours                    NA         NA      NA       NA
## among                    NA         NA      NA       NA
## non               8.501e+01  7.051e+05       0        1
## wed                      NA         NA      NA       NA
```

```
## event                    NA         NA      NA       NA
## fast                     NA         NA      NA       NA
## fire                     NA         NA      NA       NA
## connect                  NA         NA      NA       NA
## tool                     NA         NA      NA       NA
## kit                      NA         NA      NA       NA
## box                      NA         NA      NA       NA
## success                  NA         NA      NA       NA
## healthi                  NA         NA      NA       NA
## month                    NA         NA      NA       NA
## display                  NA         NA      NA       NA
## interchang               NA         NA      NA       NA
## allow_user               NA         NA      NA       NA
## theme                    NA         NA      NA       NA
## make_fashion             NA         NA      NA       NA
## water             1.169e+02  7.867e+05       0        1
## contain                  NA         NA      NA       NA
## top                      NA         NA      NA       NA
## eas                      NA         NA      NA       NA
## websit                   NA         NA      NA       NA
## peopl                    NA         NA      NA       NA
## purchas                  NA         NA      NA       NA
## partner                  NA         NA      NA       NA
## larg                     NA         NA      NA       NA
## prospect                 NA         NA      NA       NA
## transform                NA         NA      NA       NA
## carri                    NA         NA      NA       NA
## bed                      NA         NA      NA       NA
## magic                    NA         NA      NA       NA
## day                      NA         NA      NA       NA
## electr                   NA         NA      NA       NA
## right                    NA         NA      NA       NA
## music                    NA         NA      NA       NA
## real                     NA         NA      NA       NA
## favorit                  NA         NA      NA       NA
## instruct                 NA         NA      NA       NA
## open                     NA         NA      NA       NA
## point                    NA         NA      NA       NA
## serious                  NA         NA      NA       NA
## butter                   NA         NA      NA       NA
## oil                      NA         NA      NA       NA
## gluten                   NA         NA      NA       NA
## now                      NA         NA      NA       NA
## name                     NA         NA      NA       NA
## insid                    NA         NA      NA       NA
## inflat                   NA         NA      NA       NA
## chang                    NA         NA      NA       NA
## better                   NA         NA      NA       NA
## comfort                  NA         NA      NA       NA
```

```
## adjust                NA          NA          NA          NA
## varieti               NA          NA          NA          NA
## best                  NA          NA          NA          NA
## space                 NA          NA          NA          NA
## flat                  NA          NA          NA          NA
## creativ               NA          NA          NA          NA
## chocol                NA          NA          NA          NA
## corpor                NA          NA          NA          NA
## improv                NA          NA          NA          NA
## equip                 NA          NA          NA          NA
## technolog             NA          NA          NA          NA
## although              NA          NA          NA          NA
## number                NA          NA          NA          NA
## might                 NA          NA          NA          NA
## conveni               NA          NA          NA          NA
## rest                  NA          NA          NA          NA
## alway                 NA          NA          NA          NA
## rather                NA          NA          NA          NA
## smart                 NA          NA          NA          NA
## kitchen               NA          NA          NA          NA
## treat                 NA          NA          NA          NA
## cut                   NA          NA          NA          NA
## paper                 NA          NA          NA          NA
## accord                NA          NA          NA          NA
## can_use               NA          NA          NA          NA
## price                 NA          NA          NA          NA
## togeth                NA          NA          NA          NA
## everi                 NA          NA          NA          NA
## rang                  NA          NA          NA          NA
## premium_qualiti       NA          NA          NA          NA
## meal                  NA          NA          NA          NA
## portion               NA          NA          NA          NA
## wast                  NA          NA          NA          NA
## workout               NA          NA          NA          NA
## secret                NA          NA          NA          NA
## visibl                NA          NA          NA          NA
## wearer                NA          NA          NA          NA
## mother                NA          NA          NA          NA
## drink                 NA          NA          NA          NA
## put                   NA          NA          NA          NA
## thank                 NA          NA          NA          NA
## wheel                 NA          NA          NA          NA
## inch                  NA          NA          NA          NA
## nut                   NA          NA          NA          NA
## singl                 NA          NA          NA          NA
## peanut                NA          NA          NA          NA
## social                NA          NA          NA          NA
## call                  NA          NA          NA          NA
## smartphon             NA          NA          NA          NA
```

```
## build                NA          NA          NA          NA
## app                  NA          NA          NA          NA
## reusabl              NA          NA          NA          NA
## sturdi               NA          NA          NA          NA
## phone                NA          NA          NA          NA
## grow                 NA          NA          NA          NA
## afford               NA          NA          NA          NA
## content              NA          NA          NA          NA
## consum               NA          NA          NA          NA
## pay                  NA          NA          NA          NA
## may                  NA          NA          NA          NA
## feel                 NA          NA          NA          NA
## form                 NA          NA          NA          NA
## ad                   NA          NA          NA          NA
## roll                 NA          NA          NA          NA
## california           NA          NA          NA          NA
## balanc               NA          NA          NA          NA
## control              NA          NA          NA          NA
## support              NA          NA          NA          NA
## cool                 NA          NA          NA          NA
## heat                 NA          NA          NA          NA
## `function`           NA          NA          NA          NA
## remot                NA          NA          NA          NA
## tast                 NA          NA          NA          NA
## hair                 NA          NA          NA          NA
## outdoor              NA          NA          NA          NA
## innov                NA          NA          NA          NA
## invent               NA          NA          NA          NA
## love                 NA          NA          NA          NA
## sound                NA          NA          NA          NA
## snack                NA          NA          NA          NA
## someth               NA          NA          NA          NA
## result               NA          NA          NA          NA
## process              NA          NA          NA          NA
## delici               NA          NA          NA          NA
## versatil             NA          NA          NA          NA
## goe                  NA          NA          NA          NA
## simpl                NA          NA          NA          NA
## complet              NA          NA          NA          NA
## mission              NA          NA          NA          NA
## detail               NA          NA          NA          NA
## old                  NA          NA          NA          NA
## great                NA          NA          NA          NA
## audio                NA          NA          NA          NA
## promot               NA          NA          NA          NA
## iphon                NA          NA          NA          NA
## nation               NA          NA          NA          NA
## team                 NA          NA          NA          NA
## duo                  NA          NA          NA          NA
```

```
## behind                    NA         NA      NA       NA
## tablet                    NA         NA      NA       NA
## meet                      NA         NA      NA       NA
## select                    NA         NA      NA       NA
## world                     NA         NA      NA       NA
## respons                   NA         NA      NA       NA
## beach                     NA         NA      NA       NA
## typic                     NA         NA      NA       NA
## companion                 NA         NA      NA       NA
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4.8097e+02  on 346  degrees of freedom
## Residual deviance: 2.0131e-09  on   0  degrees of freedom
## AIC: 694
##
## Number of Fisher Scoring iterations: 25
```

```
#Validate the model
predTest1 = predict(LRmodel_new, newdata = test1, type="response")
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type
## == : prediction from a rank-deficient fit may be misleading
```

```
table(test1$deal, predTest1>0.3)
```

```
##
##     FALSE TRUE
##   0    42   31
##   1    43   32
```

```
(78+70)/nrow(na.omit(test1))
```

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

```
library(ROCR)
ROCRpred1 = prediction(predTest1, test1$deal)
as.numeric(performance(ROCRpred1, "auc")@y.values)
```

```
## [1] 0.5010046
```

```
perf = performance(ROCRpred1, "tpr","fpr")
plot(perf)
```