Content Based Recommendation System

Mritunjay And Sunil

13/10/2019

Recommended System Techniques on Airbnb (Amsterdam Hotel Recommendation)

1. Content Based Recommendation System

Load all the relevant libraries and Get the working directory and Load the Amsterdam Hotel Airbn data set

```
pacman::p_load(tidyverse, purrr, stringr, data.table, modelr, readxl,caret, corrplot, broom,
ggpubr, tm, proxy, MASS,relaimpo, car,interplot, caTools, mice, gbm, reshape2, compiler, reco
mmenderlab, Matrix, knitr,tidyr, dplyr, animation, wordnet, RColorBrewer, wordcloud, Snowball
C, topicmodels, ggplot2, cluster, fpc)
getwd()
```

```
## [1] "C:/Users/Rapsy/Desktop/Recommender_Assignment/MJ"
```

```
airbnb = read.csv("airbnb.csv", header=TRUE, sep=",") # transaction format!
names(airbnb) = c(colnames(airbnb))
head(airbnb,1)
```

```
##    Hotel_Id        Host_Name User_Id User_Name
## 1     2818 Erik And Mary Jo 2914515      Ivana
##                                  Hotel_name
## 1 Quiet Garden View Room & Super Fast WiFi
##                                     summary
## 1 Quiet Garden View Room & Super Fast WiFi
##
space
## 1 I'm renting a bedroom (room overlooking the garden) in my apartment in Amsterdam,  The r
oom is located to the east of the city centre in a quiet, typical Amsterdam neighbourhood the
"Indische Buurt". AmsterdamÃ¢\200\231s historic centre is less than 15 minutes away by bike o
r tram. The features of the room are: - Twin beds (80 x 200 cm, down quilts and pillows)  - 2
pure cotton towels for each guest  - reading lamps - bedside table - wardrobe - table with ch
airs - tea and coffee making facilities - mini bar - alarm clock - Hi-Fi system with cd playe
r, connection for mp3 player / phone - map of Amsterdam and public transport - Wi-Fi Internet
connection  Extra services: - Bike rental
##
description
## 1 Quiet Garden View Room & Super Fast WiFi I'm renting a bedroom (room overlooking the gar
den) in my apartment in Amsterdam,  The room is located to the east of the city centre in a q
uiet, typical Amsterdam neighbourhood the "Indische Buurt". AmsterdamÃ¢\200\231s historic cen
tre is less than 15 minutes away by bike or tram. The features of the room are: - Twin beds
(80 x 200 cm, down quilts and pillows)  - 2 pure cotton towels for each guest  - reading lamp
s - bedside table - wardrobe - table with chairs - tea and coffee making facilities - mini ba
r - alarm clock - Hi-Fi system with cd player, connection for mp3 player / phone - map of Ams
terdam and public transport - Wi-Fi Internet connection  Extra services: - Bike rental Indisc
he Buurt ("Indies Neighborhood") is a neighbourhood in the eastern portion of the city of Ams
terdam, in the Dutch province of Noord-Holland. The name dates from the early 20th century an
d is derived from the fact that the neighbourhood's streets are named after islands a
##   host_id host_name property_type   room_type accommodates
## 1 4070804    Daniel    Apartment Private room   Two Person
##            bathrooms     bedrooms    beds bed_type
## 1 One attach bathroom One bedroom One bed Real Bed
##
amenities
## 1 {Internet,Wifi,"Paid parking off premises","Buzzer/wireless intercom",Heating,Washer,"Sm
oke detector","Carbon monoxide detector","First aid kit","Safety card","Fire extinguisher",Es
sentials,Shampoo,"Lock on bedroom door","24-hour check-in",Hangers,"Hair dryer",Iron,"Laptop
friendly workspace","translation missing: en.hosting_amenity_49","translation missing: en.hos
ting_amenity_50","Private entrance","Hot water","Bed linens","Extra pillows and blankets","Si
ngle level home","Garden or backyard","No stairs or steps to enter","Flat path to guest entra
nce","Well-lit path to entrance","No stairs or steps to enter","Accessible-height bed","No st
airs or steps to enter","Host greets you","Handheld shower head","Paid parking on premises"}
##          cancellation_policy Ratings
## 1 strict_14_with_grace_period       3
```

Structure of Datasets

```
str(airbnb)
```

```
## 'data.frame':    20677 obs. of  20 variables:
##  $ Hotel_Id          : int  2818 2818 2818 2818 2818 2818 2818 2818 2818 2818 ...
##  $ Host_Name         : Factor w/ 508 levels "Aafje","Adriana",..: 136 136 136 136 136 136
136 136 136 136 ...
##  $ User_Id           : int  2914515 5711109 2944771 4620679 373226 2200958 1348274 543307
6 2847616 857406 ...
##  $ User_Name         : Factor w/ 2932 levels "(Email hidden by Airbnb)",..: 1205 1153 287
5 1130 2021 2308 413 2823 569 1964 ...
##  $ Hotel_name        : Factor w/ 507 levels "'Westerpark Sanctuary', Office-Apartmen
t",..: 383 383 383 383 383 383 383 383 383 383 ...
##  $ summary           : Factor w/ 382 levels "","'LORE'S PLACE' A lovely, open writers hom
e in the fun 'Indische Buurt' in Amsterdam! We are offering a open pla"| __truncated__,..: 24
2 242 242 242 242 242 242 242 242 242 ...
##  $ space             : Factor w/ 504 levels "","- 100 m2 floor space - private garden of
45 m2  - living room with a '30s bar, 55 inch QLED TV and home cinema "| __truncated__,..: 15
8 158 158 158 158 158 158 158 158 158 ...
##  $ description       : Factor w/ 506 levels "'LORE'S PLACE' A lovely, open writers home i
n the fun 'Indische Buurt' in Amsterdam! We are offering a open pla"| __truncated__,..: 317 3
17 317 317 317 317 317 317 317 317 ...
##  $ host_id           : int  4070804 4070804 4070804 4070804 4070804 4070804 4070804 40708
04 4070804 4070804 ...
##  $ host_name         : Factor w/ 404 levels "Aafje","Adriana",..: 81 81 81 81 81 81 81 81
81 81 ...
##  $ property_type     : Factor w/ 15 levels "Apartment","Bed and breakfast",..: 1 1 1 1 1
1 1 1 1 1 ...
##  $ room_type         : Factor w/ 3 levels "Entire home/apt",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ accommodates      : Factor w/ 10 levels "Five Person",..: 10 10 10 10 10 10 10 10 10 1
0 ...
##  $ bathrooms         : Factor w/ 11 levels "Four attach bathroom",..: 4 3 3 3 3 3 3 3 3 3
...
##  $ bedrooms          : Factor w/ 7 levels "Five bedroom",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ beds              : Factor w/ 7 levels "Five bed","Four bed",..: 3 3 3 3 3 3 3 3 3 3
...
##  $ bed_type          : Factor w/ 4 levels "Couch","Futon",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ amenities         : Factor w/ 508 levels "{\"Cable TV\",Internet,Wifi,\"Paid parking o
ff premises\",\"Buzzer/wireless intercom\",Heating,\"Family/kid fri"| __truncated__,..: 16 16
16 16 16 16 16 16 16 16 ...
##  $ cancellation_policy: Factor w/ 3 levels "flexible","moderate",..: 3 3 3 3 3 3 3 3 3 3
...
##  $ Ratings           : int  3 2 5 3 3 3 3 3 2 3 ...
```

Create a dataset for CF from main airbnb dataset (User_ID, Hotel_ID, Ratings)

```
colnames(airbnb)
```

```
##  [1] "Hotel_Id"          "Host_Name"         "User_Id"
##  [4] "User_Name"         "Hotel_name"        "summary"
##  [7] "space"             "description"       "host_id"
## [10] "host_name"         "property_type"     "room_type"
## [13] "accommodates"      "bathrooms"         "bedrooms"
## [16] "beds"              "bed_type"          "amenities"
## [19] "cancellation_policy" "Ratings"
```

Create airbnb_hotels dataframe with features columns to create corpus for text mining along with Hotel ID

```
airbnb_Hotels <- airbnb[, c('Hotel_Id', 'summary', 'space', 'description', 'property_type',
'room_type', 'accommodates', 'bathrooms', 'bedrooms', 'beds', 'bed_type', 'amenities')]
airbnb_Hotels_Unique = airbnb_Hotels[!duplicated(airbnb_Hotels$Hotel_Id), ]
head(airbnb_Hotels_Unique,4)
```

```
##      Hotel_Id
## 1       2818
## 31     20168
## 67     25428
## 108    27886
##
summary
## 1
Quiet Garden View Room & Super Fast WiFi
## 31  Cozy studio on your own private floor, 100% in a lovely 17th century Dutch townhouse i
n the heart of the city. Located a stones throw from Rembrandt Square, Dam Square, Leidse Squ
are and Flower Market. Walking distance from Central Station. Comfortable bed and with privat
e bathroom, double glazed windows and night blackout curtains, flat screen TV and mini fridg
e, Wi-Fi but no kitchen. No bathtub only shower. No curfew.
## 67
## 108
Stylish and romantic houseboat on fantastic historic location with breathtaking view. Your ow
n entrance!  Wheelhouse, deckhouse and captains room. Central, quiet. Great breakfast, 2 vanM
oof design bikes and a Canadian Canoe are included. Just read the reviews on tripadvisor for
instance!
##
space
## 1
I'm renting a bedroom (room overlooking the garden) in my apartment in Amsterdam,  The room i
s located to the east of the city centre in a quiet, typical Amsterdam neighbourhood the "Ind
ische Buurt". AmsterdamÃ¢\200\231s historic centre is less than 15 minutes away by bike or tr
am. The features of the room are: - Twin beds (80 x 200 cm, down quilts and pillows)  - 2 pur
e cotton towels for each guest  - reading lamps - bedside table - wardrobe - table with chair
s - tea and coffee making facilities - mini bar - alarm clock - Hi-Fi system with cd player,
connection for mp3 player / phone - map of Amsterdam and public transport - Wi-Fi Internet co
nnection  Extra services: - Bike rental
## 31
For those who like all facets of city life. In the streets surrounding the house, there are l
ots of antique shops, and for those who are into browsing for antiques. The Rijksmuseum is a
short walk away. The famous nine-streets with all its little boutique shops, cafes, restauran
ts and whatnot is also around the corner. The Dam Square, Leidse Square and Rembrandt Squarej
ust minutes to walk. No Curfew and you feel 100% privacy.  Hair dryer, towels and linens are
included. Airport pick-up is possible at extra charge (ask for it)
## 67  This nicely furnished, newly renovated apt  is very sunny, and spacious with 8 window
s. The appliances are fairly new. There are two flat screen TVs, washing machine, dryer, dish
washer, laptop computer, Wi-fi, and high speed internet as well as a printer. The floors are
wood throughout the apt. The large living room and bedroom are separated by a nice corridor.T
he living room has two sofas, and is very comfortable. Doors in all rooms, and lots of privac
y. Lots of books and CDs. There is a bathroom with a tub, and a separate toilet. We have a ne
w wood burning fireplace in the living room. Wonderful views of the Keizersgracht Canal and W
ester Church in front, and trees and gardens in back. The apt is in a historic area, as well
as a very convenient one. It is two blocks from the Anne Frank House. The neighborhood is fil
led with cafes, and restaurants, with  the Albery Heijn supermarket next door. There are also
neighborhood shops like a laundry, nice wine shop, bicycle rental,, etc. We a
## 108 For a romantic couple: A beautifully restored traditional dutch barge, M.s. Luctor, wi
th all modern comfords, your own entrance, three rooms,(sleeping room, deckhouse and wheelhou
se) with two bikes and a canadian canoe to explore the city. Have your breakfast weather perm
itting on the large jetty. (You  might want to give a piece of your fresh fruit to our Greek
turtle...) Organic products, a variety of cheese, fresh orange juice and croisants... We are
member of green-hotels (ECEAT) and care for the enviroment. (We just installed solar pannels!
A lot of questions can be answered on boat (Website hidden by Airbnb) there is a button with
Frequently Asked Questions.)  The location is superb, fantastic view and everything that make
```

s Amsterdam great,shops, restaurants, cafe's, red light district, musea, cinema's, theatre, i
s close by! There is wifi available, but more important: we know the city and can help you t
o find whatever you're looking for. Central station is on foot aprox. ten mi
##
description
1   Quiet Garden View Room & Super Fast WiFi I'm renting a bedroom (room overlooking the g
arden) in my apartment in Amsterdam, The room is located to the east of the city centre in a
quiet, typical Amsterdam neighbourhood the "Indische Buurt". AmsterdamÃ¢\200\231s historic ce
ntre is less than 15 minutes away by bike or tram. The features of the room are: - Twin beds
(80 x 200 cm, down quilts and pillows) - 2 pure cotton towels for each guest - reading lamp
s - bedside table - wardrobe - table with chairs - tea and coffee making facilities - mini ba
r - alarm clock - Hi-Fi system with cd player, connection for mp3 player / phone - map of Ams
terdam and public transport - Wi-Fi Internet connection Extra services: - Bike rental Indisc
he Buurt ("Indies Neighborhood") is a neighbourhood in the eastern portion of the city of Ams
terdam, in the Dutch province of Noord-Holland. The name dates from the early 20th century an
d is derived from the fact that the neighbourhood's streets are named after islands a
31   Cozy studio on your own private floor, 100% in a lovely 17th century Dutch townhouse
in the heart of the city. Located a stones throw from Rembrandt Square, Dam Square, Leidse Sq
uare and Flower Market. Walking distance from Central Station. Comfortable bed and with priva
te bathroom, double glazed windows and night blackout curtains, flat screen TV and mini fridg
e, Wi-Fi but no kitchen. No bathtub only shower. No curfew. For those who like all facets of
city life. In the streets surrounding the house, there are lots of antique shops, and for tho
se who are into browsing for antiques. The Rijksmuseum is a short walk away. The famous nine-
streets with all its little boutique shops, cafes, restaurants and whatnot is also around the
corner. The Dam Square, Leidse Square and Rembrandt Squarejust minutes to walk. No Curfew and
you feel 100% privacy. Hair dryer, towels and linens are included. Airport pick-up is possib
le at extra charge (ask for it) No curfew, free entrance 27/7 with your own
67   This nicely furnished, newly renovated apt is very sunny, and spacious with 8 windo
ws. The appliances are fairly new. There are two flat screen TVs, washing machine, dryer, dis
hwasher, laptop computer, Wi-fi, and high speed internet as well as a printer. The floors are
wood throughout the apt. The large living room and bedroom are separated by a nice corridor.T
he living room has two sofas, and is very comfortable. Doors in all rooms, and lots of privac
y. Lots of books and CDs. There is a bathroom with a tub, and a separate toilet. We have a ne
w wood burning fireplace in the living room. Wonderful views of the Keizersgracht Canal and W
ester Church in front, and trees and gardens in back. The apt is in a historic area, as well
as a very convenient one. It is two blocks from the Anne Frank House. The neighborhood is fil
led with cafes, and restaurants, with the Albery Heijn supermarket next door. There are also
neighborhood shops like a laundry, nice wine shop, bicycle rental,, etc. We a
108   Stylish and romantic houseboat on fantastic historic location with breathtaking vie
w. Your own entrance! Wheelhouse, deckhouse and captains room. Central, quiet. Great breakfa
st, 2 vanMoof design bikes and a Canadian Canoe are included. Just read the reviews on tripad
visor for instance! For a romantic couple: A beautifully restored traditional dutch barge, M.
s. Luctor, with all modern comfords, your own entrance, three rooms,(sleeping room, deckhouse
and wheelhouse) with two bikes and a canadian canoe to explore the city. Have your breakfast
weather permitting on the large jetty. (You might want to give a piece of your fresh fruit t
o our Greek turtle...) Organic products, a variety of cheese, fresh orange juice and croisant
s... We are member of green-hotels (ECEAT) and care for the enviroment. (We just installed so
lar pannels! A lot of questions can be answered on boat (Website hidden by Airbnb) there is a
button with Frequently Asked Questions.) The location is superb, fantastic vi
property_type   room_type accommodates   bathrooms
1   Apartment   Private room   Two Person One attach bathroom
31   Townhouse   Private room   Two Person One Attach bathroom
67   Apartment Entire home/apt Three person One Attach bathroom
108   Houseboat   Private room   Two Person One Attach bathroom
bedrooms   beds bed_type
1   One bedroom One bed Real Bed
31 One bedroom One bed Real Bed
67 One bedroom One bed Real Bed

```
## 108 One bedroom One bed Real Bed
##
amenities
## 1    {Internet,Wifi,"Paid parking off premises","Buzzer/wireless intercom",Heating,Washe
r,"Smoke detector","Carbon monoxide detector","First aid kit","Safety card","Fire extinguishe
r",Essentials,Shampoo,"Lock on bedroom door","24-hour check-in",Hangers,"Hair dryer",Iron,"La
ptop friendly workspace","translation missing: en.hosting_amenity_49","translation missing: e
n.hosting_amenity_50","Private entrance","Hot water","Bed linens","Extra pillows and blanket
s","Single level home","Garden or backyard","No stairs or steps to enter","Flat path to guest
entrance","Well-lit path to entrance","No stairs or steps to enter","Accessible-height be
d","No stairs or steps to enter","Host greets you","Handheld shower head","Paid parking on pr
emises"}
## 31
{TV,Internet,Wifi,"Paid parking off premises",Heating,"Smoke detector","Carbon monoxide detec
tor","Fire extinguisher",Essentials,Hangers,"Hair dryer","Laptop friendly workspace","transla
tion missing: en.hosting_amenity_49","translation missing: en.hosting_amenity_50","Hot wate
r","Bed linens",Refrigerator,"Long term stays allowed","Host greets you"}
## 67
{TV,"Cable TV",Internet,Wifi,Kitchen,Elevator,"Indoor fireplace","Buzzer/wireless intercom",H
eating,"Family/kid friendly",Washer,Dryer,"Smoke detector","Fire extinguisher",Essentials,Sha
mpoo,Hangers,"Hair dryer",Iron,"Laptop friendly workspace","Hot water","Bed linens","Extra pi
llows and blankets"}
## 108
{TV,Internet,Wifi,Breakfast,Heating,"Smoke detector","Carbon monoxide detector","Fire extingu
isher",Essentials,Shampoo,"24-hour check-in",Hangers,"Hair dryer","Laptop friendly workspac
e","translation missing: en.hosting_amenity_49","translation missing: en.hosting_amenity_5
0","Self check-in","Smart lock","Private living room","Private entrance","Hot water","Lake ac
cess"}
```

Now Combined all the features columns for mining

```
airbnb_Hotels_Unique$features_text = paste(airbnb_Hotels_Unique$summary, airbnb_Hotels_Unique
$space, airbnb_Hotels_Unique$description, airbnb_Hotels_Unique$property_type, airbnb_Hotels_U
nique$room_type, airbnb_Hotels_Unique$accommodates, airbnb_Hotels_Unique$bathrooms, airbnb_Ho
tels_Unique$bedrooms, airbnb_Hotels_Unique$beds, airbnb_Hotels_Unique$bed_type, airbnb_Hotels
_Unique$amenities)
airbnb_Hotels_mining = airbnb_Hotels_Unique[,c("Hotel_Id","features_text")]
write.csv(airbnb_Hotels_mining,"airbnb_Hotels_mining.csv")
head(airbnb_Hotels_mining,2)
```

```
##      Hotel_Id
## 1       2818
## 31     20168
##
features_text
## 1  Quiet Garden View Room & Super Fast WiFi I'm renting a bedroom (room overlooking the ga
rden) in my apartment in Amsterdam,  The room is located to the east of the city centre in a
quiet, typical Amsterdam neighbourhood the "Indische Buurt". AmsterdamÃ¢\200\231s historic ce
ntre is less than 15 minutes away by bike or tram. The features of the room are: - Twin beds
(80 x 200 cm, down quilts and pillows)  - 2 pure cotton towels for each guest  - reading lamp
s - bedside table - wardrobe - table with chairs - tea and coffee making facilities - mini ba
r - alarm clock - Hi-Fi system with cd player, connection for mp3 player / phone - map of Ams
terdam and public transport - Wi-Fi Internet connection  Extra services: - Bike rental Quiet
Garden View Room & Super Fast WiFi I'm renting a bedroom (room overlooking the garden) in my
apartment in Amsterdam,  The room is located to the east of the city centre in a quiet, typic
al Amsterdam neighbourhood the "Indische Buurt". AmsterdamÃ¢\200\231s historic centre is less
than 15 minutes away by bike or tram. The features of the room are: - Twin beds (80 x 200 cm,
down quilts and pillows)  - 2 pure cotton towels for each guest  - reading lamps - bedside ta
ble - wardrobe - table with chairs - tea and coffee making facilities - mini bar - alarm cloc
k - Hi-Fi system with cd player, connection for mp3 player / phone - map of Amsterdam and pub
lic transport - Wi-Fi Internet connection  Extra services: - Bike rental Indische Buurt ("Ind
ies Neighborhood") is a neighbourhood in the eastern portion of the city of Amsterdam, in the
Dutch province of Noord-Holland. The name dates from the early 20th century and is derived fr
om the fact that the neighbourhood's streets are named after islands a Apartment Private room
Two Person One attach bathroom One bedroom One bed Real Bed {Internet,Wifi,"Paid parking off
premises","Buzzer/wireless intercom",Heating,Washer,"Smoke detector","Carbon monoxide detecto
r","First aid kit","Safety card","Fire extinguisher",Essentials,Shampoo,"Lock on bedroom doo
r","24-hour check-in",Hangers,"Hair dryer",Iron,"Laptop friendly workspace","translation miss
ing: en.hosting_amenity_49","translation missing: en.hosting_amenity_50","Private entranc
e","Hot water","Bed linens","Extra pillows and blankets","Single level home","Garden or backy
ard","No stairs or steps to enter","Flat path to guest entrance","Well-lit path to entranc
e","No stairs or steps to enter","Accessible-height bed","No stairs or steps to enter","Host
greets you","Handheld shower head","Paid parking on premises"}
## 31
Cozy studio on your own private floor, 100% in a lovely 17th century Dutch townhouse in the h
eart of the city. Located a stones throw from Rembrandt Square, Dam Square, Leidse Square and
Flower Market. Walking distance from Central Station. Comfortable bed and with private bathro
om, double glazed windows and night blackout curtains, flat screen TV and mini fridge, Wi-Fi
but no kitchen. No bathtub only shower. No curfew. For those who like all facets of city lif
e. In the streets surrounding the house, there are lots of antique shops, and for those who a
re into browsing for antiques. The Rijksmuseum is a short walk away. The famous nine-streets
with all its little boutique shops, cafes, restaurants and whatnot is also around the corner.
The Dam Square, Leidse Square and Rembrandt Squarejust minutes to walk. No Curfew and you fee
l 100% privacy.  Hair dryer, towels and linens are included. Airport pick-up is possible at e
xtra charge (ask for it) Cozy studio on your own private floor, 100% in a lovely 17th century
Dutch townhouse in the heart of the city. Located a stones throw from Rembrandt Square, Dam S
quare, Leidse Square and Flower Market. Walking distance from Central Station. Comfortable be
d and with private bathroom, double glazed windows and night blackout curtains, flat screen T
V and mini fridge, Wi-Fi but no kitchen. No bathtub only shower. No curfew. For those who lik
e all facets of city life. In the streets surrounding the house, there are lots of antique sh
ops, and for those who are into browsing for antiques. The Rijksmuseum is a short walk away.
The famous nine-streets with all its little boutique shops, cafes, restaurants and whatnot is
also around the corner. The Dam Square, Leidse Square and Rembrandt Squarejust minutes to wal
k. No Curfew and you feel 100% privacy.  Hair dryer, towels and linens are included. Airport
pick-up is possible at extra charge (ask for it) No curfew, free entrance 27/7 with your own
Townhouse Private room Two Person One Attach bathroom One bedroom One bed Real Bed {TV,Intern
et,Wifi,"Paid parking off premises",Heating,"Smoke detector","Carbon monoxide detector","Fire
```

```
extinguisher",Essentials,Hangers,"Hair dryer","Laptop friendly workspace","translation missin
g: en.hosting_amenity_49","translation missing: en.hosting_amenity_50","Hot water","Bed linen
s",Refrigerator,"Long term stays allowed","Host greets you"}
```

Now Create Corpus on features_text

```
HotelStopWord = c(stopwords('english'), "also", "hous", "includ", "well", "around", "fuli",
"great", "lot", "apart", "famous", "minut", "amsterdam", "love", "within", "space", "away",
"nice", "friend", "need", "best", "use", "find", "miss", "public", "close", "open", "locat",
"can", "will", "just", "like")
corpus <- VCorpus(VectorSource(airbnb_Hotels_mining$features_text))

corpusNormalization <- function(corpus) {
  corpus <- tm_map(corpus, content_transformer(tolower))
  corpus <- tm_map(corpus, removeNumbers)
  corpus <- tm_map(corpus, removeWords, HotelStopWord)
  corpus <- tm_map(corpus, removePunctuation)
  corpus <- tm_map(corpus, stemDocument)
  corpus <- tm_map(corpus, removeWords, HotelStopWord)
  corpus <- tm_map(corpus, stripWhitespace)

  return(corpus)
}

corpusNorms = corpusNormalization(corpus)

dtm <- DocumentTermMatrix(corpusNorms)
dtm_ti <- weightTfIdf(dtm)
dtm_ti
```

```
## <<DocumentTermMatrix (documents: 508, terms: 5323)>>
## Non-/sparse entries: 46975/2657109
## Sparsity           : 98%
## Maximal term length: 162
## Weighting          : term frequency - inverse document frequency (normalized) (tf-idf)
```

```
mat_ti <- as.matrix(dtm_ti)

### Count the frequency of each concept
CorpusFreq <- colSums(as.matrix(dtm)) #Sum each column
head(CorpusFreq)
```

```
##     Ã¢\230â□º     Ã¢\230â□¦  Ã¢\230â□¦Ã¢\230â□¦ Ã¢\230â□¦citi Ã¢\230â□¦easi Ã¢\230â□¦fa
st
##       4       93       4       2       6       2
```

Histograms for most occured words

```
wf <- data.frame(term = names(CorpusFreq), occurrences = CorpusFreq)

ggplot(subset(wf, CorpusFreq > 250), aes(term, occurrences, color = term)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```
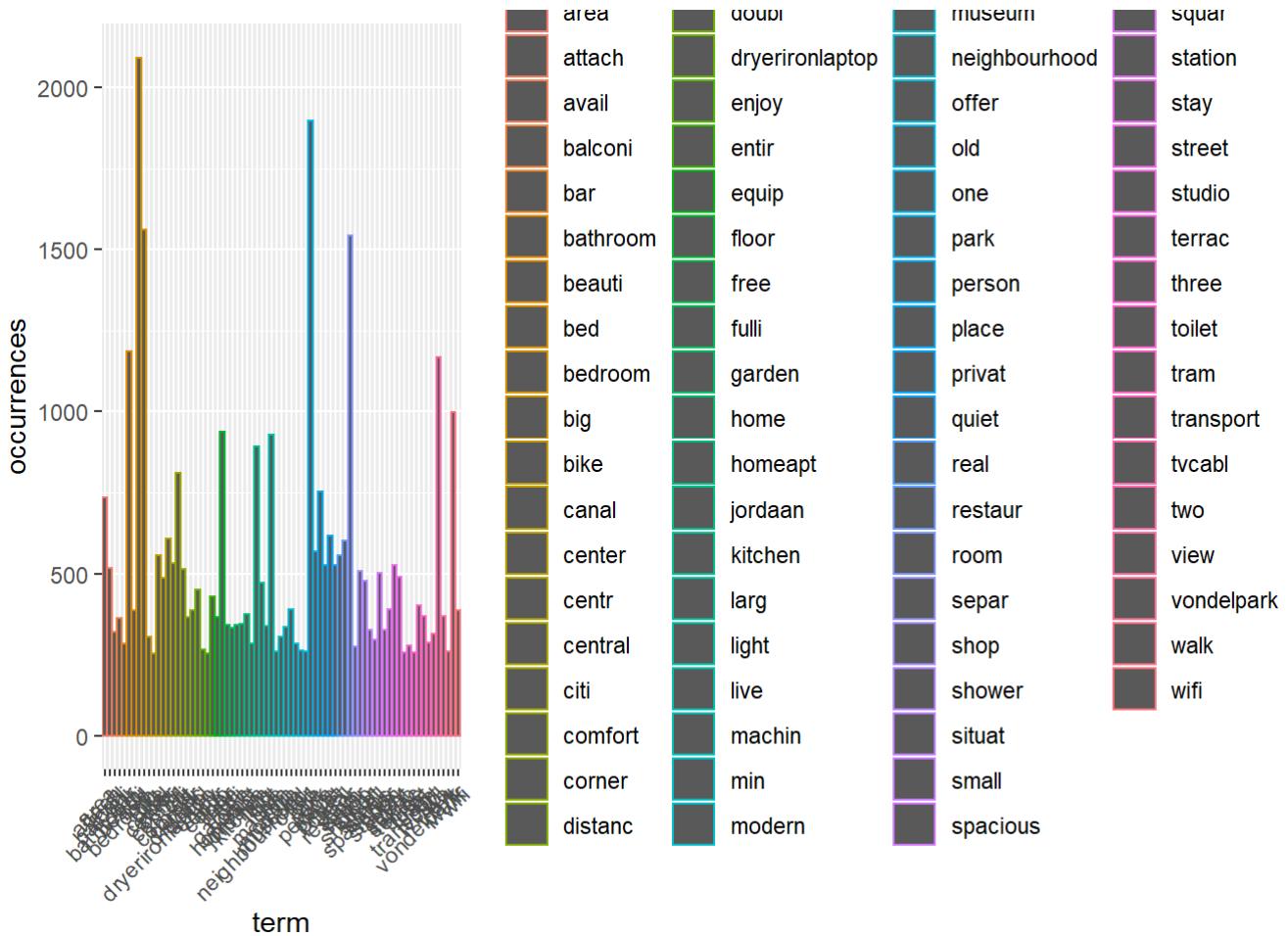

Add color on wordcloud for Accident Type Column

```
dark2 <- brewer.pal(6, "Dark2")
wordcloud(names(CorpusFreq), CorpusFreq, max.words=100, rot.per=0.6, colors=dark2)
```

Find the Frequency of ReportTypeCorpusFreq

```
length(CorpusFreq)    ### [1] 2725
```

```
## [1] 5323
```

```
CorpusOrderFreq <- order(CorpusFreq, decreasing = TRUE)

CorpusFreq[head(CorpusOrderFreq,10)]
```

```
##     bed     one bedroom    room bathroom     two    walk   floor
##    2093    1897    1561    1544    1187    1169     999     937
##    live  kitchen
##     930     891
```

```
# kill    fall   injur  struck     die    burn
# 2139    2114    1724    1151    1093     982

CorpusFreq[tail(CorpusOrderFreq)]
```

```
## xboxdvd yacuzzi    yesy    zelf   zodat    zona
##       1       1       1       1       1       1
```

```r
rownames(airbnb_Hotels_mining) <- 1:nrow(airbnb_Hotels_mining)
dist_mat_cos <- as.matrix(dist(mat_ti, method = "cosine"))
dist_mat_cos_df = data.frame(sort(dist_mat_cos[, 1], decreasing = T)[1:5])
c = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]
dist_mat_cos_df
```

```
##       sort.dist_mat_cos...1...decreasing...T..1.5.
## 489                                      0.9999959
## 422                                      0.9994104
## 430                                      0.9977646
## 496                                      0.9966062
## 340                                      0.9965599
```

```r
Top1_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top2_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top3_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top4_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top5_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))

for (i in 1:nrow(airbnb_Hotels_mining)) {
  target = airbnb_Hotels_mining[rownames(airbnb_Hotels_mining)[i],1]
  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[1])
  Top1_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[2])
  Top2_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[3])
  Top3_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[4])
  Top4_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[5])
  Top5_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]
}

df_cb <- data.frame(Top1_Similar_Hotel,
                    Top2_Similar_Hotel,
                    Top3_Similar_Hotel,
                    Top4_Similar_Hotel,
                    Top5_Similar_Hotel, stringsAsFactors = TRUE)
rownames(df_cb) = airbnb_Hotels_mining[,1]
write.csv(df_cb, file = "Content Based Recommended Hotel For each Hotel using cosine.csv")



dist_mat_cos <- as.matrix(dist(mat_ti, method = "euclidean"))
dist_mat_cos_df = data.frame(sort(dist_mat_cos[, 1], decreasing = T)[1:10])
e = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]
dist_mat_cos_df
```

```
##     sort.dist_mat_cos...1...decreasing...T..1.10.
## 305                              1.1722198
## 398                              1.1065931
## 224                              1.0310324
## 489                              1.0244151
## 422                              0.9849109
## 430                              0.8918450
## 260                              0.8395919
## 261                              0.8072879
## 340                              0.7697950
## 310                              0.7611819
```

```r
Top1_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top2_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top3_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top4_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top5_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))

for (i in 1:nrow(airbnb_Hotels_mining)) {
  target = airbnb_Hotels_mining[rownames(airbnb_Hotels_mining)[i],1]
  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[1])
  Top1_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[2])
  Top2_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[3])
  Top3_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[4])
  Top4_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[5])
  Top5_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]
}

df_cb <- data.frame(Top1_Similar_Hotel,
                    Top2_Similar_Hotel,
                    Top3_Similar_Hotel,
                    Top4_Similar_Hotel,
                    Top5_Similar_Hotel, stringsAsFactors = TRUE)
rownames(df_cb) = airbnb_Hotels_mining[,1]
write.csv(df_cb, file = "Content Based Recommended Hotel For each Hotel using euclidean.csv")

### correlation is nothing but pearson's correlation
dist_mat_cos <- as.matrix(dist(mat_ti, method = "correlation"))
dist_mat_cos_df = data.frame(sort(dist_mat_cos[, 1], decreasing = T)[1:10])
p = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]
dist_mat_cos_df
```

```
##      sort.dist_mat_cos...1...decreasing...T..1.10.
## 490                                 0.9997411
## 398                                 0.9996723
## 349                                 0.9993256
## 305                                 0.9992840
## 261                                 0.9992827
## 248                                 0.9991329
## 454                                 0.9990728
## 44                                  0.9985584
## 55                                  0.9984625
## 204                                 0.9980742
```

```
Top1_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top2_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top3_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top4_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top5_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))

for (i in 1:nrow(airbnb_Hotels_mining)) {
  target = airbnb_Hotels_mining[rownames(airbnb_Hotels_mining)[i],1]
  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[1])
  Top1_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[2])
  Top2_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[3])
  Top3_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[4])
  Top4_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[5])
  Top5_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]
}

df_cb <- data.frame(Top1_Similar_Hotel,
                    Top2_Similar_Hotel,
                    Top3_Similar_Hotel,
                    Top4_Similar_Hotel,
                    Top5_Similar_Hotel, stringsAsFactors = TRUE)
rownames(df_cb) = airbnb_Hotels_mining[,1]
write.csv(df_cb, file = "Content Based Recommended Hotel For each Hotel using correlation.cs
v")

dist_mat_cos <- as.matrix(dist(mat_ti, method = "jaccard"))
dist_mat_cos_df = data.frame(sort(dist_mat_cos[, 1], decreasing = T)[1:10])
j = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]
dist_mat_cos_df
```

```
##      sort.dist_mat_cos...1...decreasing...T..1.10.
## 489                                     0.9746193
## 204                                     0.9565217
## 496                                     0.9562842
## 173                                     0.9558011
## 422                                     0.9545455
## 196                                     0.9543147
## 305                                     0.9527027
## 210                                     0.9506173
## 277                                     0.9502762
## 349                                     0.9500000
```

```r
Top1_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top2_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top3_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top4_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))
Top5_Similar_Hotel = integer(nrow(airbnb_Hotels_mining))

for (i in 1:nrow(airbnb_Hotels_mining)) {
  target = airbnb_Hotels_mining[rownames(airbnb_Hotels_mining)[i],1]
  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[1])
  Top1_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[2])
  Top2_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[3])
  Top3_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[4])
  Top4_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]

  dist_mat_cos_df = data.frame(sort(dist_mat_cos[, i], decreasing = T)[5])
  Top5_Similar_Hotel[i] = airbnb_Hotels_mining[rownames(dist_mat_cos_df),1]
}

df_cb <- data.frame(Top1_Similar_Hotel,
                    Top2_Similar_Hotel,
                    Top3_Similar_Hotel,
                    Top4_Similar_Hotel,
                    Top5_Similar_Hotel, stringsAsFactors = TRUE)
rownames(df_cb) = airbnb_Hotels_mining[,1]
write.csv(df_cb, file = "Content Based Recommended Hotel For each Hotel using jaccard.csv")

union(union(union(c,e),p),j)
```

```
##  [1] 793845 729317 735973 800723 642190 593483 705423 481664 550783 551495
## [11] 601882 794322 652660 528227 758600  73208  91535 437757 356424 415619
## [21] 444935 570753
```

```
i1 = intersect(c,e)
i2 = intersect(c,p)
i3 = intersect(c,j)
i4 = intersect(e,p)
i5 = intersect(e,j)
i6 = intersect(p,j)
union(union(union(union(union(i1,i2),i3),i4),i5),i6)
```

```
## [1] 793845 729317 735973 642190 800723 593483 705423 551495 652660 437757
```

View