**Sentiment Analysis**

> Amazon <- read.csv("C:/Amazon.csv")

> library(tm)

> corpus<-iconv(Amazon$ReviewBody,to="utf-8")

> corpus<-Corpus(VectorSource(corpus))

> corpus<-tm\_map(corpus,tolower)

Warning message:

In tm\_map.SimpleCorpus(corpus, tolower) : transformation drops documents

> corpus<-tm\_map(corpus,removePunctuation)

Warning message:

In tm\_map.SimpleCorpus(corpus, removePunctuation) :

transformation drops documents

> inspect(corpus[1:5])

<<SimpleCorpus>>

Metadata: corpus specific: 1, document level (indexed): 0

Content: documents: 5

[1] no doubt it has a great bass and to a great extent noise cancellation and decent sound clarity and mindblowing battery but the following dissapointed me though i tried a lot to adjust1bluetooth range not more than 10m2 pain in ear due the conical budscan be removed3 wires are a bit long which makes it odd in front4 no pouch provided5 worst part is very low quality and distoring mic other person keeps complaining about my voice\n

[2] this earphones are unreliable i bought it before 15 days meanwhile right side ear buds got cracked automatically and it got divided in two parts and sound quality is also not that much good but ok one more thing bass is not good as it is a boat earphonesguysalso for the proof i have attached picsplease see and think before buying‚¬â€¹ this unreliable productthanks\n

[3] i bought itfor 999i purchased it second time gifted first one to brother this is really goodsound quality is really goodyou can connect 2 devices at a timei connect laptop and mobile bothwhen listening muaic on lappy if u get a call it switches to mobilethis switch between device is not good but again in this price its awesomebattery stand by is 36 hours for me and i have used it continously for 8 hours in one go so battery life is pretty awesomecomfortable on ear\n

[4] its sound quality is adorable overall it was good but just for 2 weeks after that it stopped working and since then it never powered on again i am claiming for warrenty and it is still undergoing\n

[5] its awesome good sound quality 89 hrs battery life just waww look ã°å¸ëœâ€° i brought just 999 rupeesbut amazons packageing system is very bad very poor packaging flipcarts packaging system is much better than amazon\n

> corpus<-tm\_map(corpus,removeNumbers)

Warning message:

In tm\_map.SimpleCorpus(corpus, removeNumbers) :

transformation drops documents

> inspect(corpus[1:5])

<<SimpleCorpus>>

Metadata: corpus specific: 1, document level (indexed): 0

Content: documents: 5

[1] no doubt it has a great bass and to a great extent noise cancellation and decent sound clarity and mindblowing battery but the following dissapointed me though i tried a lot to adjustbluetooth range not more than m pain in ear due the conical budscan be removed wires are a bit long which makes it odd in front no pouch provided worst part is very low quality and distoring mic other person keeps complaining about my voice\n

[2] this earphones are unreliable i bought it before days meanwhile right side ear buds got cracked automatically and it got divided in two parts and sound quality is also not that much good but ok one more thing bass is not good as it is a boat earphonesguysalso for the proof i have attached picsplease see and think before buyingã¢â‚¬â€¹ this unreliable productthanks\n

[3] i bought itfor i purchased it second time gifted first one to brother this is really goodsound quality is really goodyou can connect devices at a timei connect laptop and mobile bothwhen listening muaic on lappy if u get a call it switches to mobilethis switch between device is not good but again in this price its awesomebattery stand by is hours for me and i have used it continously for hours in one go so battery life is pretty awesomecomfortable on ear\n

[4] its sound quality is adorable overall it was good but just for weeks after that it stopped working and since then it never powered on again i am claiming for warrenty and it is still undergoing\n

[5] its awesome good sound quality hrs battery life just waww look ã°å¸ëœâ€° i brought just rupeesbut amazons packageing system is very bad very poor packaging flipcarts packaging system is much better than amazon\n

> corpus<-tm\_map(corpus,removeWords,stopwords('english'))

Warning message:

In tm\_map.SimpleCorpus(corpus, removeWords, stopwords("english")) :

transformation drops documents

> corpus<-tm\_map(corpus,stripWhitespace)

Warning message:

In tm\_map.SimpleCorpus(corpus, stripWhitespace) :

transformation drops documents

> tdm<-TermDocumentMatrix(corpus)

> tdm

<<TermDocumentMatrix (terms: 14748, documents: 14337)>>

Non-/sparse entries: 163840/211278236

Sparsity : 100%

Error in nchar(Terms(x), type = "chars") :

invalid multibyte string, element 49

> tdm<-as.matrix(tdm)

> install.packages("wordcloud")

Installing package into ‘C:/Users/lenovo/Documents/R/win-library/3.5’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/wordcloud\_2.6.zip'

Content type 'application/zip' length 596342 bytes (582 KB)

downloaded 582 KB

package ‘wordcloud’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\lenovo\AppData\Local\Temp\RtmpcTOFCj\downloaded\_packages

> tdm[1:6,1:2]

Docs

Terms 1 2

adjustbluetooth 1 0

bass 1 1

battery 1 0

bit 1 0

budscan 1 0

cancellation 1 0

> library(wordcloud)

Loading required package: RColorBrewer

> w<-sort(rowSums(tdm),decreasing = TRUE)

> set.seed(222)

> wordcloud(words = names(w),freq = w,max.words = 150,random.order = F,min.freq = 5,colors=brewer.pal

+

+ (8,'Dark2'))

Warning messages:

1: In wordcloud(words = names(w), freq = w, max.words = 150, random.order = F, :

charge could not be fit on page. It will not be plotted.

2: In wordcloud(words = names(w), freq = w, max.words = 150, random.order = F, :

delivery could not be fit on page. It will not be plotted.

3: In wordcloud(words = names(w), freq = w, max.words = 150, random.order = F, :

around could not be fit on page. It will not be plotted.

4: In wordcloud(words = names(w), freq = w, max.words = 150, random.order = F, :

everything could not be fit on page. It will not be plotted.

5: In wordcloud(words = names(w), freq = w, max.words = 150, random.order = F, :

buds could not be fit on page. It will not be plotted.

6: In wordcloud(words = names(w), freq = w, max.words = 150, random.order = F, :

purchased could not be fit on page. It will not be plotted.

> wordcloud(words = names(w),freq = w,max.words = 150,random.order = F,min.freq = 5,colors=brewer.pal

+

+ (8,'Dark2'),rot.per=0.3)

Warning messages:

1: In wordcloud(words = names(w), freq = w, max.words = 150, random.order = F, :

everything could not be fit on page. It will not be plotted.

2: In wordcloud(words = names(w), freq = w, max.words = 150, random.order = F, :

comes could not be fit on page. It will not be plotted.

3: In wordcloud(words = names(w), freq = w, max.words = 150, random.order = F, :

ever could not be fit on page. It will not be plotted.

4: In wordcloud(words = names(w), freq = w, max.words = 150, random.order = F, :

purchased could not be fit on page. It will not be plotted.

> Amazon <- read.csv("C:/Amazon.csv")

> Amazon<-iconv(Amazon$ReviewBody,to="utf-8")

> library(ggplot2)

Attaching package: ‘ggplot2’

The following object is masked from ‘package:NLP’:

annotate

> library(scales)

> library(reshape2)

> 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

> install.packages("syuzhet")

Installing package into ‘C:/Users/lenovo/Documents/R/win-library/3.5’

(as ‘lib’ is unspecified)

also installing the dependencies ‘textshape’, ‘dtt’

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/textshape\_1.6.0.zip'

Content type 'application/zip' length 519505 bytes (507 KB)

downloaded 507 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/dtt\_0.1-2.zip'

Content type 'application/zip' length 21279 bytes (20 KB)

downloaded 20 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/syuzhet\_1.0.4.zip'

Content type 'application/zip' length 2930157 bytes (2.8 MB)

downloaded 2.8 MB

package ‘textshape’ successfully unpacked and MD5 sums checked

package ‘dtt’ successfully unpacked and MD5 sums checked

package ‘syuzhet’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\lenovo\AppData\Local\Temp\RtmpcTOFCj\downloaded\_packages

> library(syuzhet)

Attaching package: ‘syuzhet’

The following object is masked from ‘package:scales’:

rescale

> install.packages("lubridate")

Installing package into ‘C:/Users/lenovo/Documents/R/win-library/3.5’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/lubridate\_1.7.4.zip'

Content type 'application/zip' length 1571036 bytes (1.5 MB)

downloaded 1.5 MB

package ‘lubridate’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\lenovo\AppData\Local\Temp\RtmpcTOFCj\downloaded\_packages

> library(lubridate)

Attaching package: ‘lubridate’

The following object is masked from ‘package:base’:

date

> Amazon <- read.csv("C:/Amazon.csv")

> Amazon<-iconv(Amazon$ReviewBody,to="utf-8")

> s<-get\_nrc\_sentiment(Amazon)

> head(s)

anger anticipation disgust fear joy sadness surprise trust negative positive

1 1 1 0 2 0 2 0 1 4 1

2 1 1 0 1 1 0 1 3 2 1

3 1 4 0 0 2 0 1 3 1 4

4 0 1 0 0 2 0 1 1 0 3

5 2 1 1 1 1 1 1 2 2 1

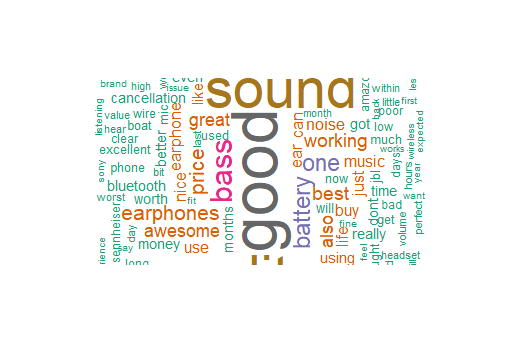
6 0 0 0 0 0 0 0 1 0 2

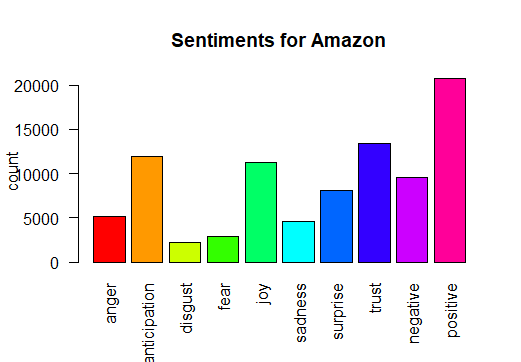
> Amazon[4]

[1] "Its sound quality is adorable. overall it was good but just for 2 weeks after that it stopped working and since then it never powered on again. I am claiming for warrenty and it is still undergoing.\n"

> barplot(colSums(s),las=2,col=rainbow(10),ylab='count',main='Sentiments for Amazon')







**Conclusion:**

The word cloud shows that the Amazon Earphone Reviews are Positive.We can conclude with following words such as good sound Awesome great.

**K-Shingles**

|  |
| --- |
| > readinteger<-function()  +  + {  +  + u1<-readLines("C:/dataset/pubg.txt")  +  + u1<-removePunctuation(u1)  +  + mystopwords=c("of","a","and","the","in","to","for","that","is","on","are","with","as","by","be","an","which","it","from","or","can","have","these","has","such")  +  + u1<-removeWords(u1,mystopwords)  +  + n<-readline(prompt="enter value of k-1:")  +  + k<-as.integer(n)  +  + Shingle<-0  +  + i<-0  +  + while(i<nchar(u1)-k+1)  +  + {  +  + Shingle[i]<-substr(u1,start = i,stop = i+k)  +  + print(Shingle[i])  +  + i=i+1  +  +  +  + }  +  +  +  + }  **Output:-**  > readinteger()  enter value of k-1:2  character(0)  [1] "Pla"  [1] "lay"  [1] "aye"  [1] "yer"  [1] "erU"  [1] "rUn"  [1] "Unk"  [1] "nkn"  [1] "kno" |
|  |
| |  | | --- | |  | |

**Time Series analysis innovative**

> mydata <- read.csv("C:/data/occupancyrates.csv")

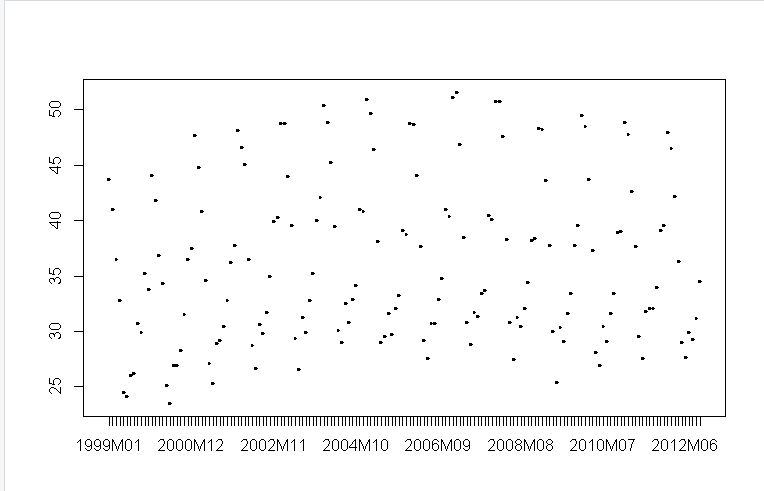
> attach(mydata)

> x<-Time

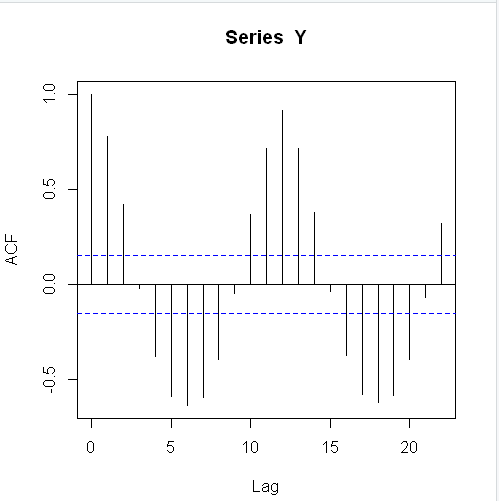
> Y<-Occupancy.rate

> d.y<-diff(Y)

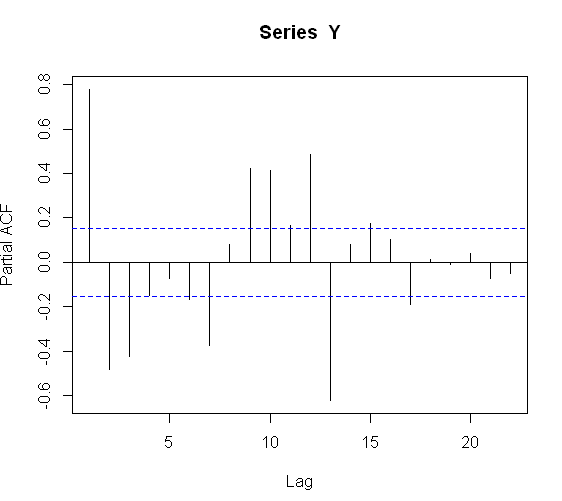
> plot(x,Y)



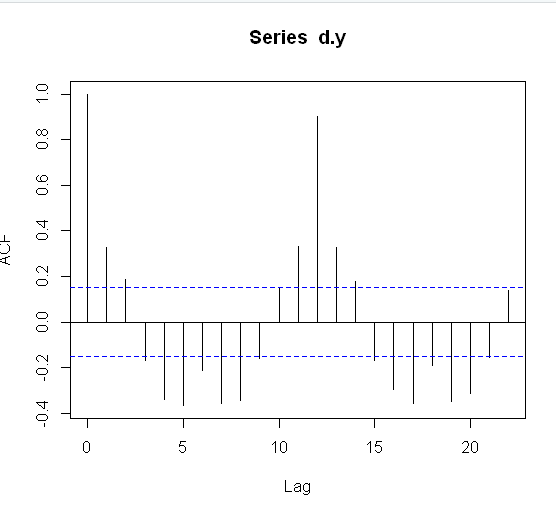
> acf(Y)



> pacf(Y)



> acf(d.y)



> arima(Y,order=c(1,0,1))

Call:

arima(x = Y, order = c(1, 0, 1))

Coefficients:

ar1 ma1 intercept

0.6868 0.3153 36.3027

s.e. 0.0625 0.0649 1.3809

sigma^2 estimated as 18.46: log likelihood = -478.13, aic = 964.26

> arima(Y,order=c(0,0,1))

Call:

arima(x = Y, order = c(0, 0, 1))

Coefficients:

ma1 intercept

0.6601 36.2179

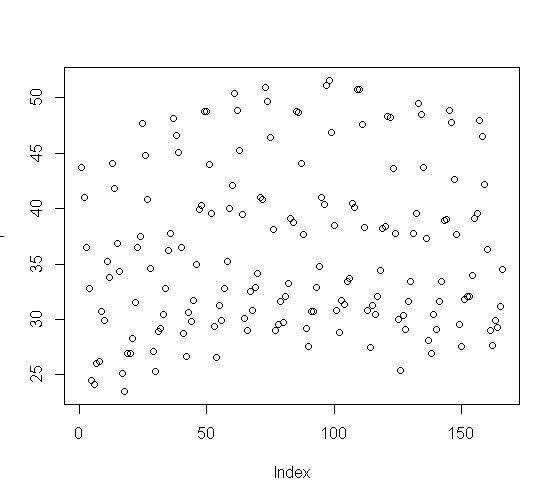
s.e. 0.0388 0.6733

sigma^2 estimated as 27.44: log likelihood = -510.72, aic = 1027.45

> mydata.arima001<-arima(Y,order =c(0,0,1) )

> mydata.pred1<-predict(mydata.arima001,n.head=001)

> plot(Y)



> lines(mydata.pred1$pred,col="blue")

> attach(mydata.pred1)

> head(mydata.pred1)

$`pred`

Time Series:

Start = 167

End = 167

Frequency = 1

[1] 35.86919

$se

Time Series:

Start = 167

End = 167

Frequency = 1

[1] 5.238236

> tail(mydata.pred1$pred)

[1] 35.86919

> head(mydata.pred1$pred)

[1] 35.86919

**Conclusion:**

In the arctic region as the time period will increase,the occupancy rate will also increases.

**Document Similarity**

> minhash<- minhash\_generator(200, seed = 235)

> ats<- TextReuseCorpus(dir="E:/rec.sport.hockey/rec.sport.hockey",n=5,minhash\_func = minhash)

Loading, tokenizing, and hashing 100 documents.

|=====================================================================================| 100%

> buckets<- lsh(ats, bands = 50 ,progress = interactive())

Calculating LSH buckets

|=====================================================================================| 100%

> candidates<- lsh\_candidates(buckets)

> my.df<- lsh\_compare(candidates, ats , jaccard\_similarity)

Making 18 comparisons.

|=====================================================================================| 100%

> my.df

# A tibble: 18 x 3

a b score

*<chr>* *<chr>* *<dbl>*

1 52558 52580 0.531

2 52558 52616 0.686

3 52571 52577 0.232

4 52571 52617 0.870

5 52577 52617 0.232

6 52580 52593 0.368

7 52580 52616 0.447

8 52582 52597 0.585

9 52586 52588 0.481

10 52586 52598 0.4

11 52588 52598 0.459

12 52589 52605 0.532

13 52589 52621 0.577

14 52589 52623 0.473

15 52605 52621 0.378

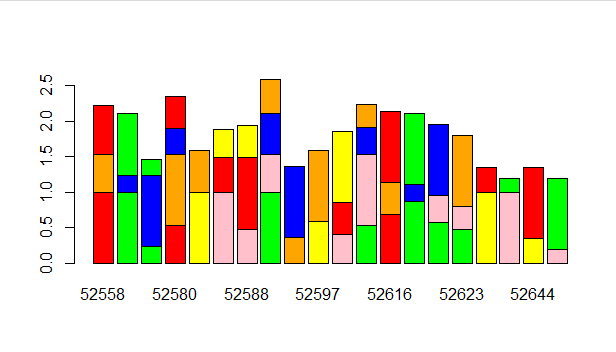
16 52605 52623 0.320

17 52625 52644 0.344

18 52638 52645 0.198

> color <- c("red" , "green" , "blue" , "orange" , "yellow" , "pink")

> barplot(as.matrix(my.df),col=color)



> ats<- TextReuseCorpus(dir="E:/rec.sport.hockey/rec.sport.hockey",n=2,minhash\_func = minhash)

Loading, tokenizing, and hashing 100 documents.

|=========================================================================| 100%

> buckets<- lsh(ats, bands = 50 ,progress = interactive())

Calculating LSH buckets

|=========================================================================| 100%

> candidates<- lsh\_candidates(buckets)

> my.df<- lsh\_compare(candidates, ats , jaccard\_similarity)

Making 26 comparisons.

|=========================================================================| 100%

> my.df

# A tibble: 26 x 3

a b score

*<chr>* *<chr>* *<dbl>*

1 52550 52567 0.262

2 52557 52576 0.273

3 52558 52580 0.559

4 52558 52593 0.349

5 52558 52616 0.731

6 52565 52601 0.177

7 52569 52615 0.342

8 52570 52609 0.205

9 52571 52617 0.899

10 52580 52593 0.430

# ... with 16 more rows

> color <- c("red" , "green" , "blue" , "orange" , "yellow" , "pink")

> barplot(as.matrix(my.df),col=color)

