1) Word Cloud – Amazon Dataset:

Clear the Ram and import the dataset:

```
> rm(list=ls())
> amazon=read.csv("amazon.csv")
```

Data Cleaning and Re-organizing:

```
> regex func=function(x){
    return (gsub('[^a-z ]','',x))
+ }
> custom_stopwords=c()
> common_stopwords=stopwords()
> all_stop_words=c(custom_stopwords,common_stopwords)
> docs=as.character(amazon$reviewText)
> docs=VCorpus(VectorSource(docs))
> docs=tm_map(docs,content_transformer(tolower))
> docs=tm_map(docs,content_transformer(regex_func))
> docs=tm_map(docs,removeWords,all_stop_words)
> docs=tm map(docs,stripWhitespace)
> dtm=DocumentTermMatrix(docs)
> df dtm=as.data.frame(as.matrix(dtm))
> dim(df dtm)
[1] 999 9322
> words_freq=colSums(df_dtm)
> library(data.table)
> words_freq=as.data.frame(words_freq)
> setDT(words_freq, keep.rownames = TRUE)[]
              rn words_freq
   1:
             aaa
                          1
   2:
                          1
             aac
   3: abandoning
                          1
9320:
                          3
         zooming
9321:
                          1
          zooms
                          2
9322:
            zune
> colnames(words freq)=c('words','frequency')
> library(dplyr)
> words freq %>% arrange(-frequency) %>% head(10)
     words frequency
1
      nook
                1454
2
      can
                 749
3
    books
                 605
  kindle
4
                 566
5
      one
                 543
                 475
6
  screen
7
                 460
      just
      like
8
                 453
9
      read
                 434
10
                 422
   great
```

Word-Cloud:

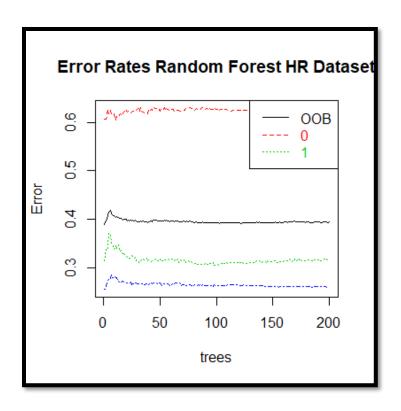
```
nothing longer interface
                                                                                    plug power real
                                                                                                                     makes display compared
                                                                                  issue tablets
                                  thought epub recommend buttons money installed
                                                pages format though doesnt light different cover devices
                                                     deel size problem able without right wanted actually
                   plus the however home already small barnes plus the however home already small barnes plus the however home already small barnes plus the however home the home back already small barnes plus the home the home back already the home back already the home barnes plus the home back already the home back already
                                                                                                                                                                                              ည္ player
                                                                                                                               really many available
                                                                                                         great
                                                                                                                                                                            install
                                                                                                                                   easy app
             looking purchased love device
                                                                                                                                                                          long features
buying pay big unit mount might found orive several Sure said on now kill
                                                                                                                                    took
                                                                                                                                                 touch
                                                                                        will like
                                                                                                                little just buy cant
                                                                                                                                                                                   text
                                                                                                                                                                got fast
                                                                            kindle
                                                                                                                                                                                            people
                                                                                                                                                       time see works works software
                                                                                                                   books
Frun inch pdf fine wall get ive android nice left get files pretty
                                                                                                                                                                                           happy nooks
                                                                                                                                                    software
 quite google library good
                                                                                                                                                  well store
                                                                                                                                                                                          day youre
                                                                                                                                                                           since
                                                                 first
                                                                                                                                                                                            hold working
                 try far
                                                                                                                                                    book thing
                                      battery
                                                                                         can of screen
                                                                                                                                                                                               either
                                                            also
                 charge Way bit
                                                                                                                                                                        another days g
          give etc made two still dont
                                                                                                                                                                    wifi ebooks
                                                        want read one use price
                                                                                                                                                                                           every
                                                                                          reading work apps need better better better mew life
          comes fire old find times
                                                                                                                                                                                          hardware
    <sub>mounting</sub> update
                           seems think bought yet tablet
                                                                                                                                                                                        support
                                                                                                                                                                                        came
                     ever never take
                                                                     cable
                                                                                                                                                                                              monitor
                                                                                                                                                     didnt set
                                around web
                                                                    ipad
                   open button hard ebook say video need
                                                                                                                              hdmi noble problems
                                                                                                              amazon
                            micro went something quality
                                                                                                                                        look B access
                                                                                                                                                                              deal
                                                                                                             download
                                        browser things full
                                                                                       know
                                                                                                                                         going phone ability
                                                               market computer everything hours worked young phone market reviews worked young phone worked young phone memory
                                                                                                              worked
                                       especially
                                                                                      games easily feature
                                                                                                                                                 extra
                                                      experience
                                                                                                       couple
                                                                                                                           havent
```

2) Sentiment Analysis on Hotstar Review Dataset:

```
Data Cleaning and Re-organizing:
> library(tm)
> library(wordcloud)
> regex func=function(x){
  return (gsub('[^a-z #@]','',x))
+ }
> custom stopwords=c()
> common stopwords=stopwords()
> all_stop_words=c(custom_stopwords,common_stopwords)
> hotstar=read.csv("hotstar.csv")
> View(hotstar)
> dim(hotstar)
[1] 5053
> docs=as.character(hotstar$Reviews)
> docs=VCorpus(VectorSource(docs))
> docs=tm_map(docs,content_transformer(tolower))
> docs=tm map(docs,content transformer(regex func))
> docs=tm_map(docs,removeWords,all_stop_words)
> docs=tm map(docs,stripWhitespace)
> dtm=DocumentTermMatrix(docs)
> df dtm=as.data.frame(as.matrix(dtm))
> dim(df dtm)
[1] 5053 7107
> words_freq=colSums(df_dtm)
> library(data.table)
> words_freq=as.data.frame(words_freq)
> setDT(words_freq, keep.rownames = TRUE)[]
                       rn words freq
            #aadhaarmemes
                                   2
                                   2
   2: #aalaporaantamizhan
   3:
                 #aarambh
                                   4
   4:
               #abeergiri
                                   2
   5: #addictionproblems
                                   1
7103:
                     zero
                                   1
7104:
                    zinab
                                   1
7105:
                 zindabad
                                   2
7106:
                     znmd
                                   1
7107:
                  zombies
                                   1
> colnames(words_freq)=c('words','frequency')
> colnames(words_freq)=c('words','frequency')
> library(dplyr)
> top_words=words_freq %>% arrange(-frequency) %>% head(25)
> df_dtm_subset=df_dtm[,top_words$words]
> df dtm subset$sentiment=hotstar$Sentiment Manual
> dim(df dtm subset)
[1] 5053
           26
```

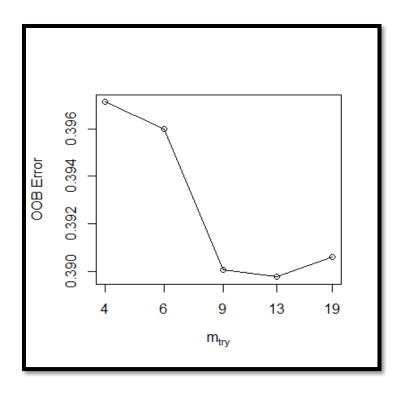
Splitting the Dataset into Test and Train:

```
> newdata=sample(1:nrow(df_dtm_subset),size=0.3*nrow(df_dtm_subset))
> test = df_dtm_subset[newdata, ]
> train = df dtm subset[-newdata,]
> dim(test)
[1] 1515 26
> dim(train)
[1] 3538 26
> attach(train)
Random Forest Modeling:
> library(randomForest)
> RF <- randomForest(as.factor(sentiment) ~ ., data = train, ntree=200, mtry = 6,
nodesize = 100, importance=TRUE)
> print(RF)
Call:
mtry = 6, nodesize = 100, importance = TRUE)
             Type of random forest: classification
                   Number of trees: 200
No. of variables tried at each split: 6
       OOB estimate of error rate: 39.4%
Confusion matrix:
        Negative Neutral Positive class.error
Negative
            413
                   343 354 0.6279279
Neutral
            123
                    839
                            263
                                 0.3151020
Positive
            108
                    203
                            892 0.2585204
> plot(RF, main="")
> legend("topright", c("00B", "0", "1"), text.col=1:6, lty=1:3, col=1:3)
> title(main="Error Rates Random Forest HR Dataset")
```



Fine Tuning the RF Model:

```
> tRF <- tuneRF(x = train[,-26], y=as.factor(train$sentiment), mtryStart = 6,</pre>
ntreeTry=100, stepFactor = 1.5,
                improve = 0.0001, trace=TRUE, plot = TRUE, doBest = TRUE, nodesize =
100, importance=TRUE)
mtry = 6 00B error = 39.6%
Searching left ...
mtry = 4
                 00B error = 39.71\%
-0.002855103 1e-04
Searching right ...
mtry = 9
                 00B error = 39.01\%
0.01498929 1e-04
mtry = 13
                 00B error = 38.98\%
0.0007246377 1e-04
mtry = 19
                 00B error = 39.06\%
-0.002175489 1e-04
```



> tRF

Call:

randomForest(x = x, y = y, mtry = res[which.min(res[, 2]), 1], nodesize = 100, importance = TRUE)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 13

OOB estimate of error rate: 38.84%

Confusion matrix:

Negative Neutral Positive class.error Negative 453 319 338 0.5918919 Neutral 250 142 833 0.3200000 878 0.2701579 Positive 123 202

Checking the Model accuracy with Development dataset:

```
> train$predict.class <- predict(tRF, train, type="class")</pre>
```

- > train\$predict.score <- predict(tRF, train, type="prob")</pre>
- > with(train, table(sentiment, predict.class))

predict.class

p. carcercrass			
sentiment	Negative	Neutral	Positive
Negative	486	305	319
Neutral	121	857	247
Positive	118	201	884

Checking the Model accuracy with Holdout dataset:

Model Accuracy:

Train Dataset:

> (486+857+884)/(3538) [1] 0.6294517 = 62.94 %

Test Dataset:

> (200+333+385)/(1515) [1] 0.6059406 = 60.59 %