```r
rm(list =ls(all = TRUE))
setwd("C:/Users/bvkka/Desktop/edureka")
library(readr)
RestaurantReviews<-read_tsv(file = 'M8_Restaurant_Reviews.tsv', na=":")

#install.packages("tm")
#install.packages("SnowballC")
library(tm)
library(SnowballC)
library(dplyr)



# Creating the corpus

corpusRR<- Corpus(VectorSource(RestaurantReviews))

##Perform the following text cleaning techniques

corpusRR <- Corpus(VectorSource(RestaurantReviews))%>%
  tm_map(removePunctuation)%>%
  tm_map(removeNumbers)%>%
  tm_map(content_transformer(tolower))%>%
  tm_map(removeWords, stopwords("english"))%>%
  tm_map(stripWhitespace)%>%
  tm_map(stemDocument)
dtmRR <- DocumentTermMatrix(corpusRR)

dtmRR
dtmRR <- removeSparseTerms(dtmRR,0.99)

dtmRR

library(wordcloud)
wordcloud(corpusRR, max.words = 40,
          random.order = FALSE,
          random.color = TRUE)
```

```r
###TASK2###


word.freqRR <- sort(colSums(as.matrix(dtmRR)),decreasing = T)
head(word.freqRR)

table.RR <- data.frame(word = names(word.freqRR),
                       absolute.frequency = word.freqRR,
                       relative.frequency = word.freqRR/length(word.freqRR))

head(table.RR)

library(ggplot2)
subset(corpusRR, frequency > 100)%>%
  ggplot(aes(word, frequency))+
  geom_bar(stat = "identity")+
  theme(axis.text.x = element_text(angle = 45))



# Save the Table as
write.csv(table.RR[1:1000,],"RR.csv")

finalTable <- table.RR %>%
  merge(table.RR, by ="word") %>%
  mutate(dProp = relative.frequency.x - relative.frequency.y,dAbs = abs(dProp)) %>%
  arrange(desc(dAbs)) %>%
  rename(RR.freq = absolute.frequency.x,RR.prop = relative.frequency.x)

head(finalTable)


###splitting the data into train and test sets


set.seed(12345)
di <- sample(2, nrow(RestaurantReviews), prob = c(0.7,0.3), replace = TRUE)
```

```r
train <- RestaurantReviews[di==1,]
test <- RestaurantReviews[di==2,]

##converting it into data frame
dtm_matrix<-as.data.frame(as.matrix(dtmRR))
View(dtm_matrix)
class(dtm_matrix)

dtm_matrix$like<-as.factor(RestaurantReviews$Liked)
##Building a classifier


#####Naive Bayes#####



library(e1071)
NB_model <- naiveBayes(Liked~., data = train)

NB_model
summary(NB_model)

test1<-as.factor(as.integer(test$Liked))

pred_NB_model <- predict (NB_model, newdata = test)

pred_NB_model



library(caret)
confusionMatrix(table(pred_NB_model, test$Liked))
```

Environment | **History** | **Connections**

Import Dataset ▾ | ≡ List ▾ | C

Global Environment ▾

Data

● corpusRR	List of 2	🔍
● dtmRR	List of 6	🔍
● Restauran...	1000 obs. of 2 variabl...	🔲

	Review	Liked
574	He also came back to check on us regularly, excellent service.	1
575	Our server was super nice and checked on us many times.	1
576	The pizza tasted old, super chewy in not a good way.	0
577	I swung in to give them a try but was deeply disappointed.	0
578	Service was good and the company was better!	1
579	The staff are also very friendly and efficient.	1
580	As for the service: I'm a fan, because it's quick and you're bei...	1
581	Boy was that sucker dry!!.	0
582	Over rated.	0
583	If you look for authentic Thai food, go else where.	0
584	Their steaks are 100% recommended!	1
585	After I pulled up my car I waited for another 15 minutes bef...	0
586	Great food and great service in a clean and friendly setting.	1
587	All in all, I can assure you I'll be back.	1
588	I hate those things as much as cheap quality black olives.	0
589	My breakfast was perpared great, with a beautiful presentati...	1
590	The kids play area is NASTY!	0
591	Great place fo take out or eat in.	1
592	The waitress was friendly and happy to accomodate for vega...	1
593	OMG I felt like I had never eaten Thai food until this dish.	1
594	It was extremely "crumby" and pretty tasteless.	0
595	It was a pale color instead of nice and char and has NO flavor.	0
596	The croutons also taste homemade which is an extra plus.	1
597	I got home to see the driest damn wings ever!	0

Files | **Plots** | **Packages** | **Help** | **Viewer**

Export ▾

```
> word.freqRR <- sort(colSums(as.matrix(dtmRR)),decreasing = T)
> head(word.freqRR)
  food  place   good servic  great   back
   125    112     95     84     70     61
```

```
> table.RR <- data.frame(word = names(word.freqRR),
+                        absolute.frequency = word.freqRR,
+                        relative.frequency = word.freqRR/length(word.freq
RR))
> head(table.RR)
         word absolute.frequency relative.frequency
food     food                125         0.07921420
place   place                112         0.07097592
good     good                 95         0.06020279
servic servic                 84         0.05323194
great   great                 70         0.04435995
back     back                 61         0.03865653
```

```
Y   My husband said she was very rude... did not even apologize for the bad
 food or anything.
 0
       0.002777778
   Review
Y   My salad had a bland vinegrette on the baby greens and hearts of Palm.
 0                                                              0.002777778
   Review
Y   My sashimi was poor quality being soggy and tasteless.
 0                                              0.002777778
   Review
Y   My side Greek salad with the Greek dressing was so tasty, and the pita
 and hummus was very refreshing.
 0
                        0.000000000
   Review
Y   Nargile - I think you are great.
 0                      0.000000000
   Review
Y   Needless to say, I won't be going back anytime soon.
 0                                      0.002777778
   Review
Y   Never again will I be dining at this place!
 0                                  0.002777778
   Review
Y   Never been to Hard Rock Casino before, WILL NEVER EVER STEP FORWARD IN
 IT AGAIN!
 0                                                                     0.
002777778
   Review
Y   Never going back. Nice ambiance.
 0        0.002777778    0.000000000
   Review
Y   Nice blanket of moz over top but i feel like this was done to cover up
 the subpar food.
 0
    0.000000000
   Review
Y   Nice, spicy and tender.
 0            0.000000000
   Review
Y   Nicest Chinese restaurant I've been in a while.
 0                                      0.000000000
   Review
Y   No allergy warnings on the menu, and the waitress had absolutely no clu
 e as to which meals did or did not contain peanuts.
```