Amazon Product Reviews - Sentiment Analysis and Fake Genuine Review Analysis

Niharika Sharma

November 20, 2019

Executive Summary

The Objective of this study is to learn how different statistical NLP techniques can be used to analyze the Consumer Behaviour(Sentiment Analysis) and to check whether the consumer is real(Authenticity of Consumer Review Analysis).

My objective is not to classify each review as either positive or negative as I am more focussed on learning the different sentiments involved in the reviews for a product. This is why I have not used Supervised Machine Learning Algorithms.

There are two parts of this Study - 1. Sentiment Analysis 2. Fake Review Analysis

Sentiment Analysis focusses on determining whether the product has positive or negative or mixed reviews in general. It does not focus on defining each review as either negative or positive as I believe that the Sentiment Analysis is more beneficial if we try to analyze the consumer sentiments associated with a product and not focus on ananlyzing whether a product review was posotive or negative.

Fake Review Analysis focusses more on analysizing consumer behaviour when it comes to writing reviews. From a business stand point, identifying fake users is more beneficial rather than identifying fake reviews because we can remove or block the users so that no more fake reviews get added to any products.

Data Upload

```
pacman::p_load(e1071, tidyverse, caret, rmarkdown, corrplot, readxl, ModelMetrics,quanteda,tidytext ,dpl
theme_set(theme_classic())
```

```
initial_data <- read_csv("Data_Amazon.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   dateAdded = col_datetime(format = ""),
##   dateUpdated = col_datetime(format = ""),
##   reviews.date = col_datetime(format = ""),
##   reviews.dateAdded = col_logical(),
##   reviews.doRecommend = col_logical(),
##   reviews.id = col_double(),
##   reviews.numHelpful = col_double(),
##   reviews.rating = col_double()
## )
```

```
## See spec(...) for full column specifications.
## Warning: 1052 parsing failures.
##  row               col           expected               actual              file
## 3424 reviews.dateAdded 1/0/T/F/TRUE/FALSE 2017-05-21T05:55:29Z 'Data_Amazon.csv'
## 3425 reviews.dateAdded 1/0/T/F/TRUE/FALSE 2017-05-21T06:00:34Z 'Data_Amazon.csv'
## 3426 reviews.dateAdded 1/0/T/F/TRUE/FALSE 2017-05-21T06:01:25Z 'Data_Amazon.csv'
## 3427 reviews.dateAdded 1/0/T/F/TRUE/FALSE 2017-03-10T06:27:42Z 'Data_Amazon.csv'
## 3428 reviews.dateAdded 1/0/T/F/TRUE/FALSE 2017-05-21T05:55:23Z 'Data_Amazon.csv'
## .... ................. .................. .................... .................
## See problems(...) for more details.
```

Popularity of Amazon Products, in General

```r
head(data.frame(sort(table(initial_data$reviews.rating),decreasing = TRUE)))
```

```
##   Var1 Freq
## 1    5 3478
## 2    4 1208
## 3    3  197
## 4    1   63
## 5    2   54
```

As we can see that most of the products fall under Satisfied(Rating - 4& 5) and Neutral(Rating - 3) so we can say that Amazon Products have a good reputation in general.

Top 5 Most Bought Amazon Products

```r
head(data.frame(sort(table(initial_data$name),decreasing = TRUE)))
```

```
##                                                                                        Var1
## 1                          Amazon Echo Show Alexa-enabled Bluetooth Speaker with 7" Screen
## 2 All-New Fire HD 8 Tablet, 8" HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta
## 3                                              Amazon - Echo Plus w/ Built-In Hub - Silver
## 4                   Fire Kids Edition Tablet, 7 Display, Wi-Fi, 16 GB, Blue Kid-Proof Case
## 5                  Brand New Amazon Kindle Fire 16gb 7" Ips Display Tablet Wifi 16 Gb Blue
## 6                    Fire Tablet, 7 Display, Wi-Fi, 16 GB - Includes Special Offers, Black
##   Freq
## 1  845
## 2  797
## 3  590
## 4  561
## 5  467
## 6  371
```

Checking Ratings for most bought product - Amazon Echo Show Alexa - 7" Screen

```r
alexa7 <- filter(initial_data,name=='Amazon Echo Show Alexa-enabled Bluetooth Speaker with 7" Screen')

data.frame(sort(table(alexa7$reviews.rating),decreasing = TRUE))
```

```
##   Var1 Freq
## 1    5  627
## 2    4  172
## 3    3   31
## 4    2    9
## 5    1    6
```

Most of the people are very satisfied with 7 inch Echo

5 Least Bought Amazon Products

```r
head(data.frame(sort(table(initial_data$name))))
```

```
##
## 1                                 Amazon Fire TV with 4K Ultra HD and Alexa Voice Remote (Pendant De
## 2                                                                                Amazon - Kindle Voy
## 3                              Amazon 9W PowerFast Official OEM USB Charger and Power Adapter for Fi
## 4 Kindle Oasis E-reader with Leather Charging Cover - Merlot, 6 High-Resolution Display (300 ppi), Wi
## 5 Kindle Oasis E-reader with Leather Charging Cover - Black, 6" High-Resolution Display (300 ppi), Wi
## 6                                         All-New Fire HD 8 Tablet, 8" HD Display, Wi-Fi, 32 GB - In
##   Freq
## 1    4
## 2   22
## 3   22
## 4   24
## 5   39
## 6   40
```

Checking Ratings for least bought product - Amazon Fire TV 4K

```r
FireTV <- filter(initial_data,name=='Amazon Fire TV with 4K Ultra HD and Alexa Voice Remote (Pendant De

data.frame(sort(table(FireTV$reviews.rating),decreasing = TRUE))
```

```
##   sort.table.FireTV.reviews.rating...decreasing...TRUE.
## 5                                                     4
```

Fire TV is the least bought product but it still has a rating of 5

Topic Modeling for Echo Alexa 7 inch

```
dtm <- dfm(alexa7$reviews.text, remove_punct=TRUE, tolower=TRUE, remove=stopwords("english"))


topic_dtm <- convert(dtm, to="topicmodels")

set.seed(1)
lda_model<- topicmodels::LDA(topic_dtm,method = "Gibbs", k=5)
terms(lda_model,5)
```

```
##      Topic 1   Topic 2  Topic 3  Topic 4  Topic 5
## [1,] "great"   "can"    "device" "echo"   "alexa"
## [2,] "love"    "screen" "good"   "show"   "music"
## [3,] "use"     "like"   "works"  "amazon" "home"
## [4,] "product" "video"  "will"   "one"    "smart"
## [5,] "bought"  "see"    "fun"    "things" "just"
```

Bar Plot of most used words for Echo Alexa 7-inch

```
# corpus conversion of the dataset
corpus_alexa <- Corpus(VectorSource(alexa7$reviews.text))
#inspect(corpus_alexa)

# cleaning the corpus
corpus_alexa <- tm_map(corpus_alexa,tolower)
```

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

```
corpus_alexa <- tm_map(corpus_alexa,removePunctuation)
```

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

```
#inspect(corpus_alexa)

# removes stopwords
cleanset <- tm_map(corpus_alexa,removeWords,stopwords('english'))
```

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

```
# remove the obviously common words
cleanset <- tm_map(cleanset, removeWords, c('alexa','devices','anyone','amazon','echo'))
```

```
## Warning in tm_map.SimpleCorpus(cleanset, removeWords, c("alexa",
## "devices", : transformation drops documents
```

```r
# removes extra whitespace
cleanset <- tm_map(cleanset, stripWhitespace)
```

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

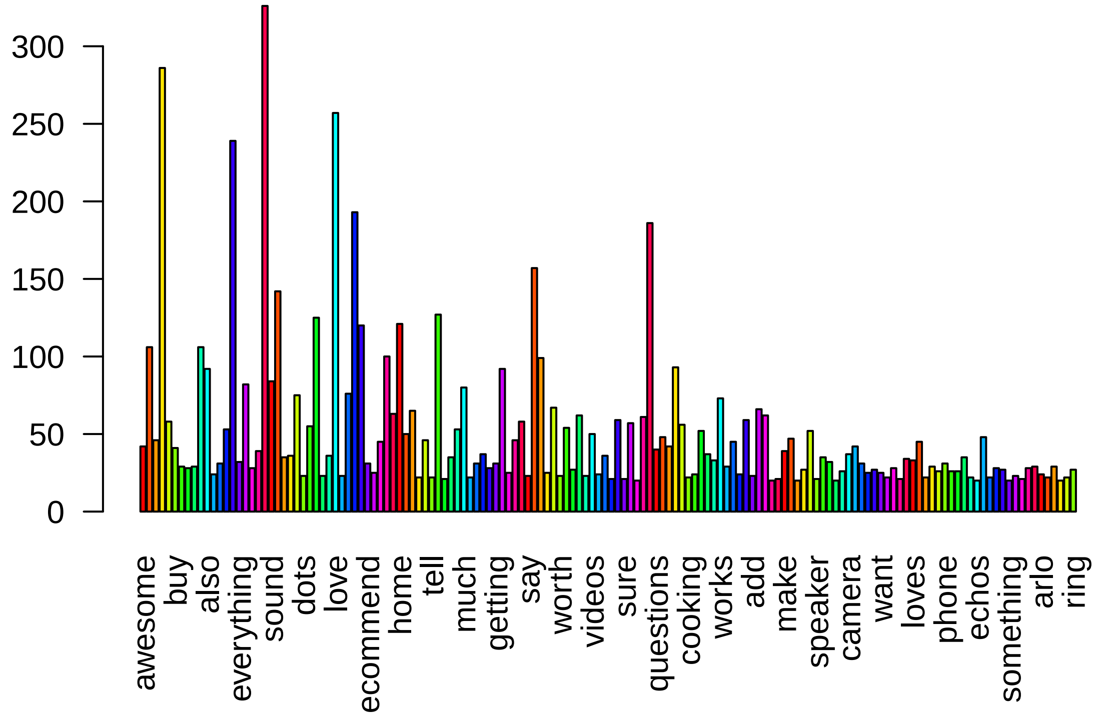
```r
# creating a term document matrix
tdm <- DocumentTermMatrix(cleanset)
tdm <- as.matrix(tdm)
tdm[1:10,1:10]
```

```
##     Terms
## Docs awesome coexist easy gift great iot setup best buy convenient
##   1        1       1    1    1     1   1     1    0   0          0
##   2        0       0    0    1     0   0     0    1   1          1
##   3        0       0    0    0     0   0     0    0   0          0
##   4        0       0    0    0     0   0     0    0   0          0
##   5        1       0    0    0     0   0     0    0   0          0
##   6        0       0    0    0     0   0     0    0   0          0
##   7        0       0    1    0     1   0     0    0   0          0
##   8        0       0    0    0     0   0     0    0   0          0
##   9        0       0    0    0     0   0     0    0   0          0
##   10       1       0    0    0     0   0     0    0   0          0
```

```r
# How many times does a word appear in term document matrix
count <- colSums(tdm)
count <- subset(count, count>=20)
count
```

```
##      awesome        easy        gift       great        best         buy
##           42         106          46         286          58          41
##     products    purchase     another      bought        also         app
##           29          28          29         106          92          24
##        calls     cameras         can  everything        good        plus
##           31          53         239          32          82          28
##     security        show       sound       video       voice       watch
##           39         326          84         142          35          36
##         well        dots         now         one        cool         lot
##           75          23          55         125          23          36
##         love     looking         set         use     product   recommend
##          257          23          76         193         120          31
##       useful         got      device         get        home      really
##           25          45         100          63         121          50
##        smart         ‚äôm   purchased        tell        like    speakers
##           65          22          46          22         127          21
##         used         dot        much       shows         two        call
##           35          53          80          22          31          37
##        every     getting        just      little        news        nice
##           28          31          92          25          46          58
##          say      screen         see        talk      things       worth
##           23         157          99          25          67          23
##          new       think        will       price      videos         etc
##          54          27          62          23          50          24
##      lyrics        need        play        sure         fun information
##          36          21          59          21          57          20
##     kitchen       music   questions     weather   christmas      family
##          61         186          40          48          42          93
##       house     cooking     display     recipes     quality       using
##          56          22          24          52          37          33
##       works        item    features    learning        time         add
##          73          29          45          24          59          23
##        able      better       enjoy         far        make        work
##          66          62          20          21          39          47
##       songs     playing        ‚äôs     speaker     youtube        many
##          20          27          52          21          35          32
##       since        wife      camera         ask         day      listen
##          20          26          37          42          31          25
##      lights        want        keep        view    shopping       still
##          27          25          22          28          21          34
##       loves    addition        chat     picture     feature       phone
##          33          45          22          29          26          31
##       prime      system     amazing        door       echos        even
##          26          26          35          22          20          48
##        help        kids       first   something        read       touch
##          22          28          27          20          23          21
##        room       thing        arlo    doorbell        drop     helpful
##          28          29          24          22          29          20
##         way        ring
##          22          27
```

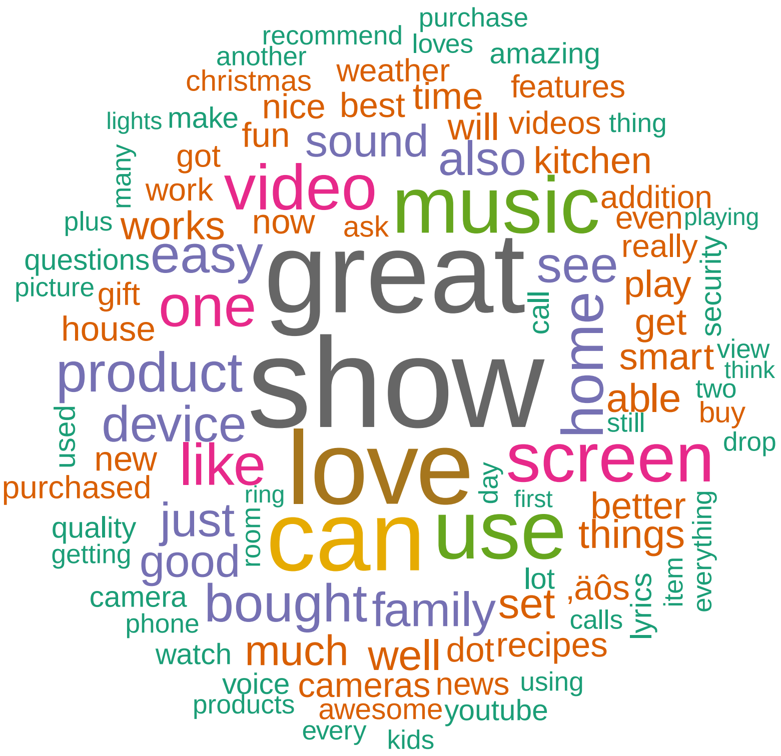
```
barplot(count, las=2, col=rainbow(20))
```



Word Cloud for Echo Alexa 7-inch

```
count <- sort(count, decreasing = TRUE)
set.seed(1)

wordcloud(words = names(count), freq = count, max.words = 100, random.order = F, min.freq = 5, colors =
```

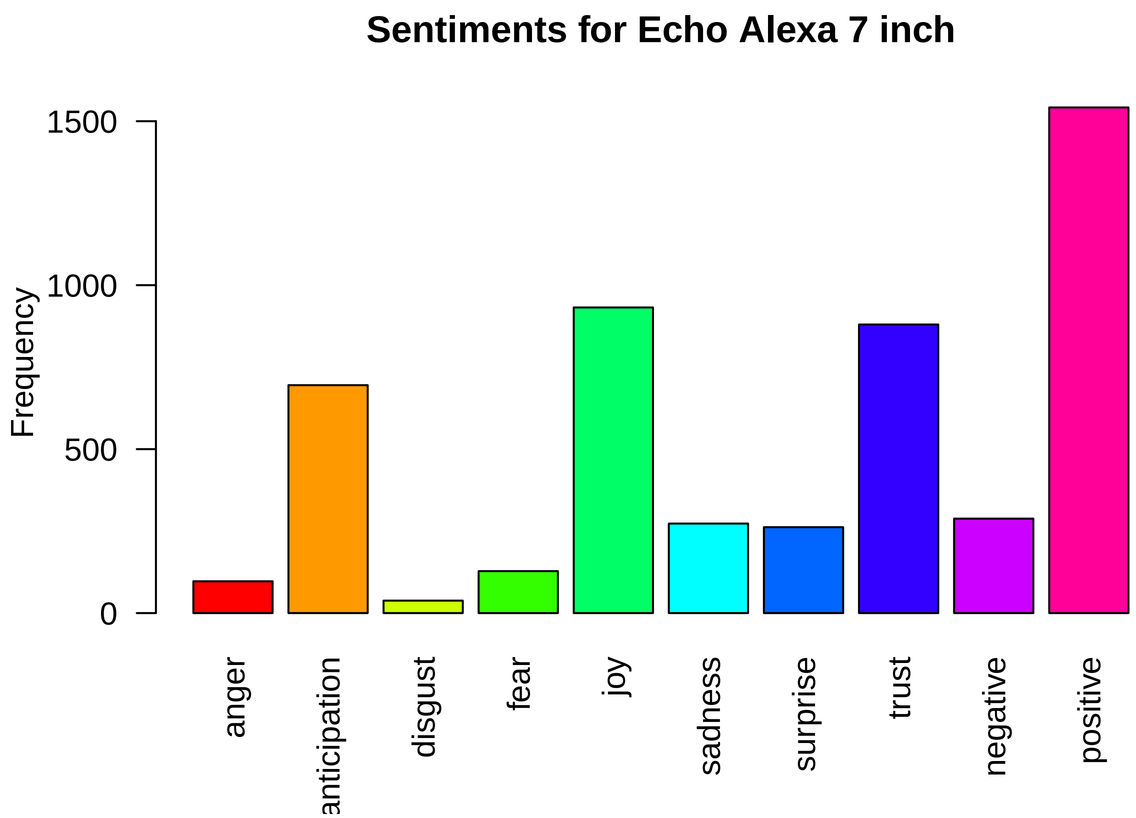


Sentiment Analysis for Most Bought Product - Amazon Echo Alexa 7 inch

```
alexa7_text <- iconv(alexa7$reviews.text, to="utf-8")
s <- get_nrc_sentiment(alexa7_text)
head(s)
```

```
##   anger anticipation disgust fear joy sadness surprise trust negative
## 1     0            1       0    0   1       0        1     1        0
## 2     0            3       0    0   2       0        2     1        0
## 3     0            1       0    0   1       1        0     2        1
## 4     1            4       0    1   2       0        1     4        2
## 5     0            0       0    0   0       0        0     1        0
## 6     0            0       0    0   1       0        0     1        0
##   positive
## 1        2
## 2        4
## 3        1
## 4        3
## 5        0
## 6        2
```

```
barplot(colSums(s), las=2, col=rainbow(10), ylab="Frequency", main="Sentiments for Echo Alexa 7 inch")
```



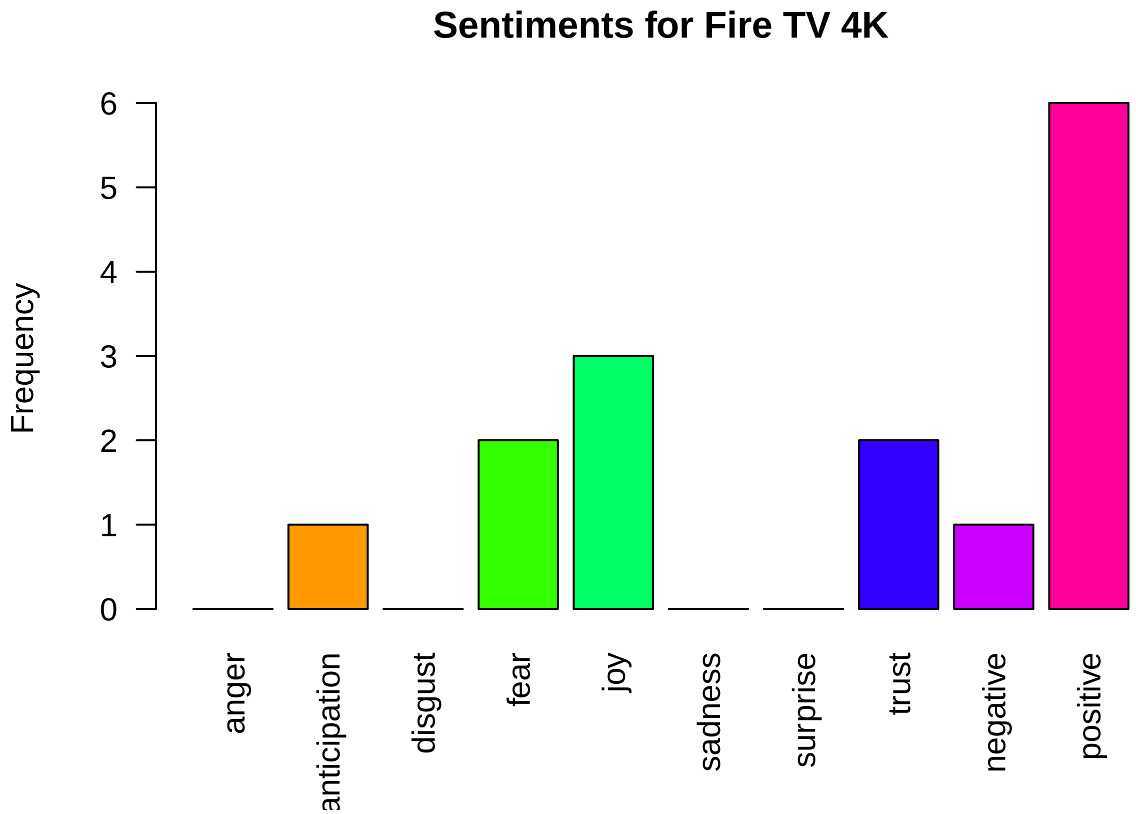
The reviews for Amazon Echo Alexa 7-inch remain positive in general.

Sentiment Analysis for Least Bought Product - Amazon Fire TV 4K

```
FireTV_text <- iconv(FireTV$reviews.text, to="utf-8")
s <- get_nrc_sentiment(FireTV_text)
head(s)
```

```
##   anger anticipation disgust fear joy sadness surprise trust negative
## 1     0            0       0    0   1       0        0     1        1
## 2     0            1       0    1   1       0        0     1        0
## 3     0            0       0    1   0       0        0     0        0
## 4     0            0       0    0   1       0        0     0        0
##   positive
## 1        2
## 2        1
## 3        1
## 4        2
```

```
barplot(colSums(s), las=2, col=rainbow(10), ylab="Frequency", main="Sentiments for Fire TV 4K")
```

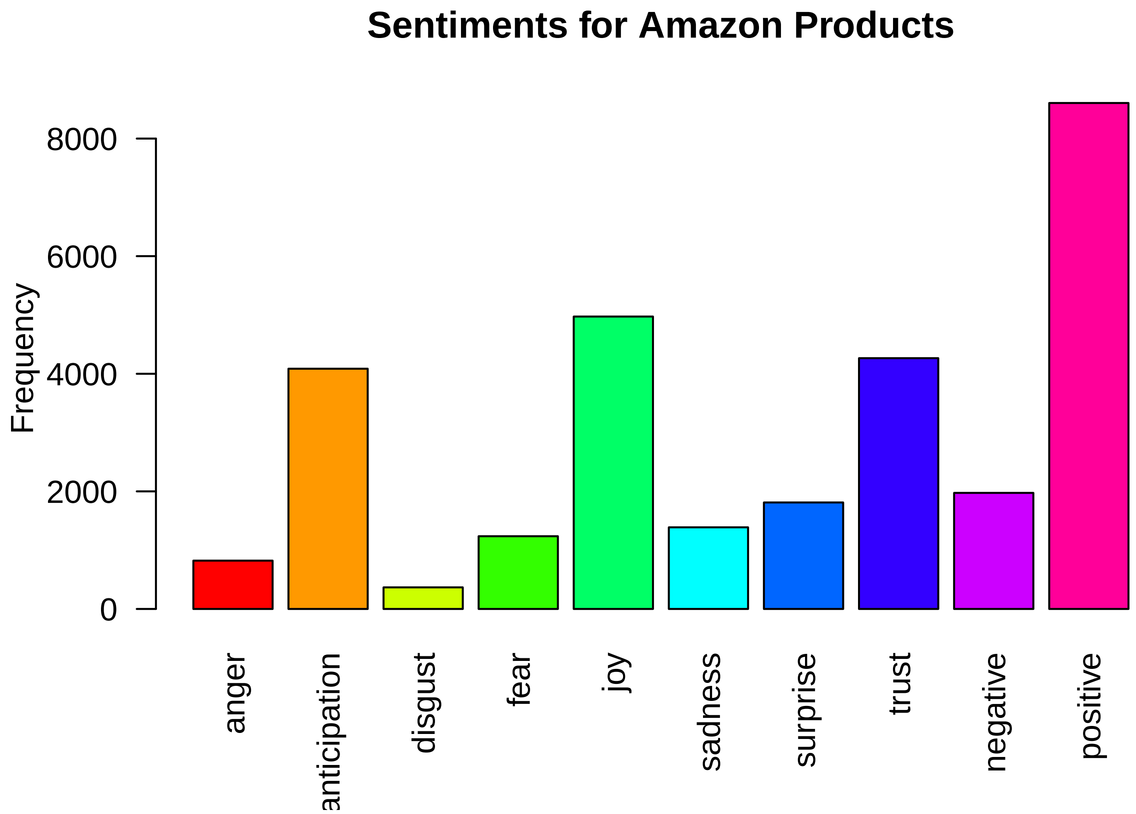


Sentiment Analysis for All the Amazon Products in general

```r
reviews_text <- iconv(initial_data$reviews.text, to="utf-8")
s <- get_nrc_sentiment(reviews_text)
head(s)
```

```
##   anger anticipation disgust fear joy sadness surprise trust negative
## 1     0            1       0    0   0       0        0     1        1
## 2     0            0       0    0   1       0        0     0        0
## 3     0            1       0    0   1       2        0     1        1
## 4     2            5       1    1   4       1        2     4        2
## 5     1            1       0    0   3       0        1     3        1
## 6     0            3       0    0   3       0        3     4        1
##   positive
## 1        1
## 2        0
## 3        1
## 4        8
## 5        7
## 6        5
```

```r
barplot(colSums(s), las=2, col=rainbow(10), ylab="Frequency", main="Sentiments for Amazon Products")
```



Sentiments Analysis Conclusion:

1. We were right in anticipating that the Amazon Products in general are very good.
2. Both the most bought product - Amazon Echo Alexa 7 inch and the least bought product - Fire TV 4K have positive sentiments associated as per the text reviews.

Extra Notes :

Problem of differentiating between "good" and "not good"

```
bing <- get_sentiments("bing")

sentiment("I am very good at drawing.")
```

```
##    element_id sentence_id word_count sentiment
## 1:          1           1          6 0.5511352
```

```
sentiment("I am not very good at drawing.")
```

```
##    element_id sentence_id word_count   sentiment
## 1:          1           1          7 -0.05669467
```

As we can see that it gives a positive result for a sentence with “good” and a negative result for the sentence having “not good” in its contents. Using the library ‘sentimentr’ in such cases can solve the problem where “not” is considered a neutral word and not a negative word.

To get an average score for the sentence.

```
sentiment_by('I am very good at drawing.', by = NULL)
```

```
##    element_id word_count sd ave_sentiment
## 1:          1          6 NA     0.5511352
```

```
sentiment_by('I am not very good at drawing.', by = NULL)
```

```
##    element_id word_count sd ave_sentiment
## 1:          1          7 NA   -0.05669467
```

‘sentiment_by’ calculated the average score for the whole sentence. This helps when we want to understand the tone of the sentence and the word count in a sentence shouldn’t have much impact. We can say that it scales the sentiment of the sentence.

Fake/Genuine Review Analysis

We will first take a look at all the people who have the most number of reviews in the data, i.e, the people having more than 10 reviews for Amazon Products.

The people who have bought these many products and written these many reviews are either:

1. Influencers - They need to buy a lot of products and genuinely provide reviews on Social Media and ecommerce websites.
2. People who buy a lot of gifts.
3. Chatbots writing a number of reviews in a loop using one username at a time.

```
head(initial_data)
```

```
## # A tibble: 6 x 24
##   id    dateAdded           dateUpdated         name  asins brand
##   <chr> <dttm>              <dttm>              <chr> <chr> <chr>
## 1 AVqV~ 2017-03-03 16:56:05 2018-10-25 16:36:31 "Ama~ B00Z~ Amaz~
## 2 AVqV~ 2017-03-03 16:56:05 2018-10-25 16:36:31 "Ama~ B00Z~ Amaz~
## 3 AVqV~ 2017-03-03 16:56:05 2018-10-25 16:36:31 "Ama~ B00Z~ Amaz~
## 4 AVqV~ 2017-03-03 16:56:05 2018-10-25 16:36:31 "Ama~ B00Z~ Amaz~
## 5 AVqV~ 2017-03-03 16:56:05 2018-10-25 16:36:31 "Ama~ B00Z~ Amaz~
## 6 AVqV~ 2017-03-03 16:56:05 2018-10-25 16:36:31 "Ama~ B00Z~ Amaz~
## # ... with 18 more variables: categories <chr>, primaryCategories <chr>,
## #   imageURLs <chr>, keys <chr>, manufacturer <chr>,
## #   manufacturerNumber <chr>, reviews.date <dttm>,
## #   reviews.dateAdded <lgl>, reviews.dateSeen <chr>,
## #   reviews.doRecommend <lgl>, reviews.id <dbl>, reviews.numHelpful <dbl>,
## #   reviews.rating <dbl>, reviews.sourceURLs <chr>, reviews.text <chr>,
## #   reviews.title <chr>, reviews.username <chr>, sourceURLs <chr>
```

```
head(data.frame(sort(table(initial_data$reviews.username),decreasing = TRUE)),10)
```

```
##         Var1 Freq
## 1       Mike   26
## 2      Chris   14
## 3       Dave   13
## 4       John   13
## 5       Nick   13
## 6       Rick   13
## 7       Bill   12
## 8     Robert   12
## 9  Anonymous   10
## 10      Brad   10
```

```
fake_data <- subset(initial_data, reviews.username %in% names(which(table(reviews.username) >= 10)))

sort(table(fake_data$reviews.username),decreasing = TRUE)
```

```
## 
##      Mike     Chris      Dave      John      Nick      Rick      Bill 
##        26        14        13        13        13        13        12 
##    Robert Anonymous      Brad     Steve      Tony 
##        12        10        10        10        10 
```

```
# Checking for how many distinct products, the username has written a review

a <-  fake_data %>%
      group_by(reviews.username) %>%
      summarise(n_distinct(name))

a
```

```
## # A tibble: 12 x 2
##    reviews.username `n_distinct(name)`
##    <chr>                         <int>
##  1 Anonymous                         7
##  2 Bill                              7
##  3 Brad                              8
##  4 Chris                             9
##  5 Dave                              7
##  6 John                              9
##  7 Mike                              9
##  8 Nick                              5
##  9 Rick                              7
## 10 Robert                            8
## 11 Steve                             7
## 12 Tony                              6
```

These people have written a lot of reviews for different unique products.

Let us check how many unique products are there in this data.

```
fake_data%>%summarise(n_distinct(name))
```

```
## # A tibble: 1 x 1
##   `n_distinct(name)`
##                <int>
## 1                 21
```

There are 21 unique products in the dataset

```
head(data.frame(sort(table(fake_data$name),decreasing = TRUE)))
```

```
##                                                                                          Var1
## 1                          Amazon Echo Show Alexa-enabled Bluetooth Speaker with 7" Screen
## 2                                                      Amazon - Echo Plus w/ Built-In Hub - Silver
## 3 All-New Fire HD 8 Tablet, 8" HD Display, Wi-Fi, 16 GB - Includes Special Offers, Magenta
## 4                    Fire Tablet, 7 Display, Wi-Fi, 16 GB - Includes Special Offers, Black
## 5                  Brand New Amazon Kindle Fire 16gb 7" Ips Display Tablet Wifi 16 Gb Blue
## 6                   Fire Kids Edition Tablet, 7 Display, Wi-Fi, 16 GB, Blue Kid-Proof Case
##   Freq
## 1   32
## 2   22
## 3   21
## 4   13
## 5   11
## 6   11
```

```
by_name_uname <- fake_data %>% group_by(reviews.username,name)
by_un_n <- by_name_uname %>% summarise(n = n())
by_un_n %>% arrange(desc(n))
```

```
## # A tibble: 89 x 3
## # Groups:   reviews.username [12]
##    reviews.username name                                                  n
##    <chr>            <chr>                                             <int>
##  1 Mike             "Amazon Echo Show Alexa-enabled Bluetooth Speake~     8
##  2 Mike             Amazon - Echo Plus w/ Built-In Hub - Silver           5
##  3 Nick             "Amazon Echo Show Alexa-enabled Bluetooth Speake~     5
##  4 Bill             Amazon - Echo Plus w/ Built-In Hub - Silver           4
##  5 Chris            "Amazon Echo Show Alexa-enabled Bluetooth Speake~     4
##  6 Mike             "All-New Fire HD 8 Tablet, 8\" HD Display, Wi-Fi~     4
##  7 Rick             Amazon - Echo Plus w/ Built-In Hub - Silver           4
##  8 Bill             Kindle E-reader - White, 6 Glare-Free Touchscree~     3
##  9 Dave             "All-New Fire HD 8 Tablet, 8\" HD Display, Wi-Fi~     3
## 10 Dave             Fire Kids Edition Tablet, 7 Display, Wi-Fi, 16 G~     3
## # ... with 79 more rows
```

As we can see that most of the users have provided more than one reviews for the same type of the Amazon Product. Even if the user bought the product more than once, there might not be a need to provide a review again as the product is listed as a different product even if the minutest details such as the color or the size is different.

As the user “Mike” has provided a lot of reviews for two different types of Amazon Echo, we will take a look at the reviews from “Mike” for these two products - “Amazon Echo Show Alexa-enabled Bluetooth Speaker with 7” Screen" and “Amazon - Echo Plus w/ Built-In Hub - Silver”.

We are check the different sentiments used by the same user “Mike” for the same product.

```
Echo_Mike <- filter(fake_data,reviews.username=='Mike')

Echo1_Mike <- filter(Echo_Mike,name=='Amazon Echo Show Alexa-enabled Bluetooth Speaker with 7" Screen')

Echo2_Mike <- filter(Echo_Mike,name=='Amazon - Echo Plus w/ Built-In Hub - Silver')
```

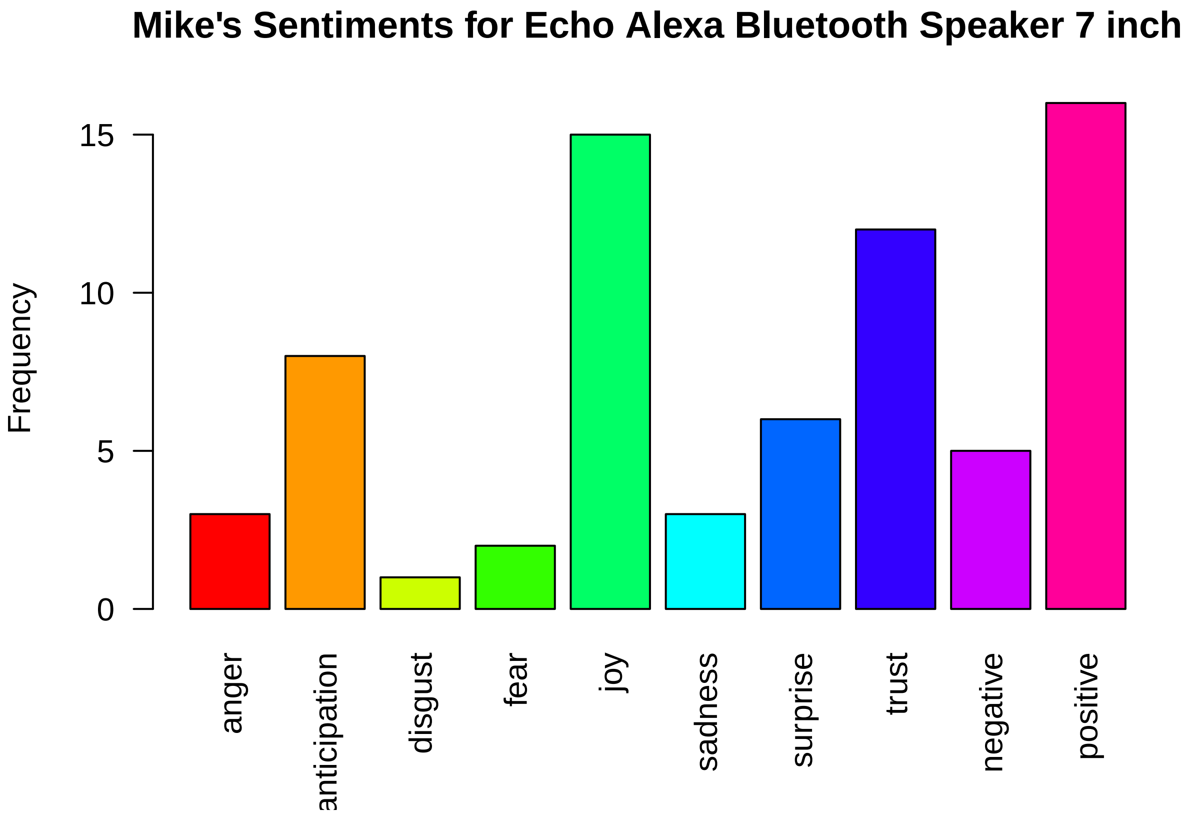
This is full of both negative and positive emotions.

Sentiments for the reviews made by Mike for Amazon Echo Show Alexa-enabled Bluetooth Speaker with 7

```
EM1 <- iconv(Echo1_Mike$reviews.text, to="utf-8")
s <- get_nrc_sentiment(EM1)
head(s)
```

```
##   anger anticipation disgust fear joy sadness surprise trust negative
## 1     0            1       0    0   1       0        1     0        0
## 2     0            0       0    0   1       0        0     0        0
## 3     0            1       0    0   3       1        1     3        1
## 4     0            1       0    0   3       1        1     3        1
## 5     3            1       1    2   1       1        1     1        3
## 6     0            0       0    0   1       0        0     0        0
##   positive
## 1        1
## 2        1
## 3        3
## 4        3
## 5        1
## 6        1
```

```
barplot(colSums(s), las=2, col=rainbow(10), ylab="Frequency", main="Mike's Sentiments for Echo Alexa Bl
```

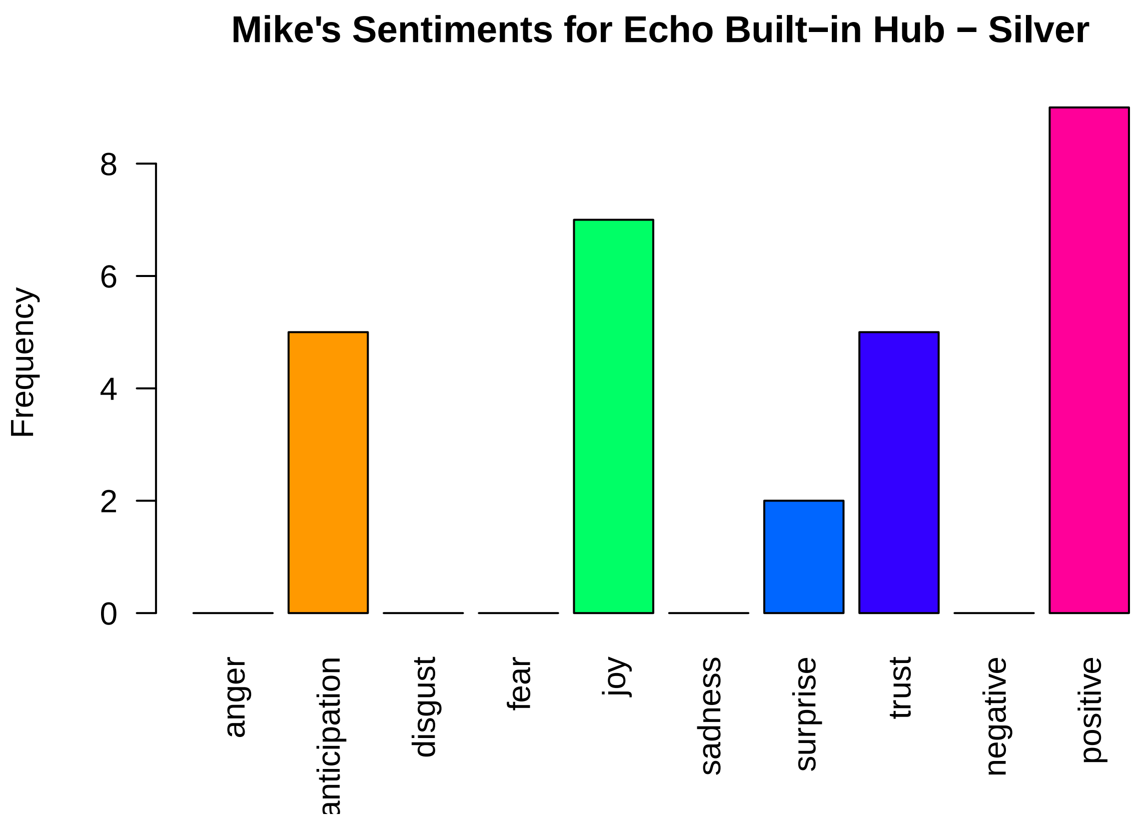


Sentiments for the reviews made by Mike for Amazon - Echo Plus w/ Built-In Hub - Silver

```
EM2 <- iconv(Echo2_Mike$reviews.text, to="utf-8")
s <- get_nrc_sentiment(EM2)
head(s)
```

```
##   anger anticipation disgust fear joy sadness surprise trust negative
## 1     0            0       0    0   0       0        0     0        0
## 2     0            0       0    0   2       0        0     1        0
## 3     0            2       0    0   2       0        1     1        0
## 4     0            2       0    0   2       0        1     3        0
## 5     0            1       0    0   1       0        0     0        0
##   positive
## 1        1
## 2        2
## 3        2
## 4        3
## 5        1
```

```
barplot(colSums(s), las=2, col=rainbow(10), ylab="Frequency", main="Mike's Sentiments for Echo Built-in
```



THis is full of only positive reviews.

Let's explore the dataset further.

```
table(Echo1_Mike$reviews.text,Echo1_Mike$dateAdded)
```

```
##
##
##   I bought this as a gift. As far as know it works fine
##   I had the Echo previously, but having the screen opens up more options for Alexa. I find it helpfu
##   So having Alexa on hand could be good and bad. It's nice when I'm cooking and I need help with a r
##   This was a Black Friday special and we knew we wanted a smart home assistant and Alexa worked well
##   We bought two of these for the kids. They love them and I now have an echo dot
##
##
##   I bought this as a gift. As far as know it works fine
##   I had the Echo previously, but having the screen opens up more options for Alexa. I find it helpfu
##   So having Alexa on hand could be good and bad. It's nice when I'm cooking and I need help with a r
##   This was a Black Friday special and we knew we wanted a smart home assistant and Alexa worked well
##   We bought two of these for the kids. They love them and I now have an echo dot
```

```
sort(table(Echo1_Mike$reviews.text),decreasing = TRUE)
```

```
##
##
##
```

```
##
##
##
##
## I had the Echo previously, but having the screen opens up more options for Alexa. I find it helpful
##
##
##
```

```
sort(table(Echo1_Mike$dateAdded),decreasing = TRUE)
```

```
##
## 2018-04-13 02:03:15 2018-02-02 02:30:22
##                   5                   3
```

After looking at the data for the Echo 7 inch for Mike's reviews, we see that most of the reviews were updated on the same dates and also most of the reviews are duplicates. And most of the ratings are a 4 or a 5. So we can say that Mike might be either a bot or is gaining some incentive by posting so many reviews. Influencers usually provide one very detailed review and don't spam the product with reviews.

Conclusion of Fake Reviews Analysis

The Objective of this Project is not to classify the reviews as fake or genuine but to identify the behavioural patterns of users which might be posting fake reviews. It is very difficult to classify one review independently as either fake or not. It is more beneficial to track one username because if it is identified as a bot then it is useful not just for one product but can be helpful for other products as well.

Fin