

# Amazon Echo reviews analysis

Elena Costagliola [211506]

12/07/2021

## 1 Introduction

The last decades have been characterized by an exponential technological progress impacting not only on people's way of living but also on how people interact and communicate. Everything has become over-personalized, from advertisement to the way our devices interact with us and vice versa. Until a few decades ago, people watched films like Star Wars, imagining that human-robot interaction was just a figment of their imagination. Today's world is far cry from the one depicted by G. Lucas, but especially for the youngest generation it is now natural to rely on virtual assistants to manage their routines, from answering simple tasks to controlling the home smartly by interconnecting more and more devices. While this kind of advancement started with virtual assistants on mobiles, in 2014 Amazon proposed Alexa, the first virtual assistant integrated into a smart speaker. This report, based on the reviews left on the Amazon Shop after the purchase of Amazon Echo devices, explores the uses and criticisms of customers and leads us to understand how these devices have changed their habits and, consequently, what features were most appreciated for each model. In the first part an explorative analysis on ratings and the models' purchased has been conducted, then opinion mining tools have been used to highlight what Alexa is used for and what is the perception the users have about it.

## 2 General Framework

This story has its roots in the 1950s, with Alan Turing's paper "Computing Machinery and Intelligence". From that moment on, humans began to imagine the world as a place of interaction between man and machine, and traces of these dreams can be found in the cinema as well as in the literature, but also in the technological advancements of that time. Today we talk about the *Internet of Things* and *smart environment*, which include smart manufacturing, smart cities and smart homes. Therefore, it is no longer a dream to be able to tell your smart device to turn the light on, or even to have your coffee ready in the morning.

This is made possible by virtual assistants (VA), that is software services that process user's textual, visual or vocal inputs, identify its meaning and respond to the command.

The first VA released on the market was Siri by Apple in 2011, followed by Google Now by Google and Cortana by Microsoft. They were all mobile or PC integrated.

Amazon was the first to integrate its virtual assistant Alexa into a smart speaker: Amazon Echo. Released in 2014, Echo is a hub for potentially all the houses' devices. It connects with its owner's Amazon account and from year to year it responds to more and more needs and with increased effectiveness. Today the smart speaker market is significantly growing as many companies are investing in this field, but Amazon Echo is still one of the most purchased products with 70% of the market share (ReaserchandMarkets, 2021). Indeed, the Echo line has grown as well: it proposes a wide range of smart speakers responding to different needs. However, like any product that achieves high popularity in a relatively short time, issues about how reliable it is have emerged both in public discussion and as subject of articles. Indeed, the voice is listened by Alexa and converted to text through Automatic Speech Recognition Alexa system; both the recorded voice and text format are saved in cloud and then sent to Natural Language Understanding system to perform syntax, semantic and pragmatic analyses. These processes respectively understand the structure, the words' meaning and the context of the question. At this point, information to answer the question are retrieved in text format from multiple sources. The user, indeed, adds the so-called skills to its device, that is applications developed by third-parties' companies and Amazon itself, which provides different services. This mechanism, which is common to many of the smart speakers on the market, has raised concerns among consumers about privacy issues. Over the years, several rumors have emerged about who has access to Alexa's recordings and how the device is constantly listening in on household conversations. Despite this, however, the market does not seem to be slowing down its growth and is expected to reach \$17.85 billion in 2025 at a compound annual growth rate (CAGR) of 26% (ReaserchandMarkets, 2021).

Two research questions emerge from these considerations:

- What are the Amazon Echo products used for?
- What is the customers' perception of Amazon Echo? Are they concerned about their privacy?

### 3 Methodology

This analysis has been conducted using RStudio, an Integrated Development Environment for R, one of the most used programming languages for statistical computing and graphics. R indeed provides many libraries aimed to conduct data manipulation, visualization and opinion mining, pivotal for the purposes of this analysis. The libraries used are listed below, roughly distinguishing their use:

```
# Data Manipulation and Visualization  
library(dplyr)  
library(tidyr)  
library(ggplot2)
```

```
# Opinion mining
library(quanteda)
library(quanteda.textstats)
library(quanteda.textplots)
library(textstem)
library(tidytext)
library(topicmodels)
library(tibble)
```

### 3.1 The Dataset

In order to answer the research questions, reviews of Amazon customers will be analysed. [Source's dataset](#) did not specify how data were collected, but they have presumably been extracted from [amazon.com](#) and refer to USA Amazon Echo buyers. Reviews from Amazon, indeed, can be extracted either through web scraping or by using third-party Application Programs Interfaces (APIs). The two methods interact with Amazon's contents allowing their extraction, but the first option may be subject to legal issues. On the one hand, indeed, the General Data Protection Regulation (GDPR) imposes limits on the processing of personal data, which means that after scraping the page, reviews authors must be anonymised. On the other hand, according to Amazon's conditions of use, a written consent of Amazon must be obtained before data extraction.

However, data have been read in R specifying encoding, columns' name and substituting empty slots with NA.

```
df <- read.csv("amazon_alexa.tsv", sep='\t', encoding="UTF-8", na.strings=c(" ", "NA"),
               col.names=c("product_rating", "review_date", "product_variation",
                           "review_text", "feedback"))
```

The dataset contains 3150 rows, each corresponding to each customer's:

- *Star rating* ranging from 1 to 5 and referred to the purchased product,
- *Date of review's publication*,
- *Model* of the product purchased within the Amazon Echo line,
- *Review's text*, and
- *Feedback*.

The summary below shows that appropriate data types have to be set to the attributes.

```
## product_rating review_date product_variation review_text
## Min. :1.000 Length:3150 Length:3150 Length:3150
```

```
## 1st Qu.:4.000   Class :character   Class :character   Class :character
## Median :5.000   Mode  :character   Mode  :character   Mode  :character
## Mean    :4.463
## 3rd Qu.:5.000
## Max.    :5.000
##      feedback
## Min.    :0.0000
## 1st Qu.:1.0000
## Median :1.0000
## Mean    :0.9184
## 3rd Qu.:1.0000
## Max.    :1.0000
```

## 3.2 Data Cleaning

In this section data will be cleaned in order to be used for the analysis using the grammar **dplyr** alongside the library **tidyr**, specific to tidy messy data. To start, the attribute **feedback** has been removed as no variable description was available. Moreover, this dataset contains 2.5% of missing values only in the attribute **review\_text**. Since no imputation is applicable to this attribute, those observations can be safely dropped by losing only 79. The attributes **product\_variation** and **product\_rating** have been converted into factors to better identify their levels and the **review\_date** has been converted into date format.

```
df <- df %>%
  select(-feedback) %>%                                # remove feedback column
  drop_na() %>%                                          # remove NAs' rows
  mutate_at(vars(product_variation, product_rating),    # set data types
             list(~factor(.))) %>%
  mutate(review_date=as.Date(review_date, format="%d-%b-%y"))
```

It has also been noted that the attribute referred to Amazon Echo variations contains 16, but two of them, specifically **Black** and **White** are ambiguous: only the color is specified, omitting the model. Reading some of their reviews, it emerged that these levels refer to the Echo Dot model, so those levels have been renamed. In addition, the one referred to the product **Fire TV Stick** has been renamed with a more appropriated name too.

```
df <- df %>% mutate(product_variation=recode(product_variation,
                                             "White"="White Dot", "Black"="Black Dot",
                                             "Configuration: Fire TV Stick"="Fire TV Stick"))
```

After these manipulations, the dataset consists of 3071 rows and 4 columns.

### 3.3 Data Exploration

As mentioned above, the dataset contains the reviews of 3071 Amazon Echo buyers. From the summary below it emerges that the reviews refer only to the period May-July 2018 and that more than twice of the customers gave a 5 rating to their Echo.

```
summary(df)
```

```
## product_rating review_date product_variation
## 1: 146          Min.    :2018-05-16 Black Dot      :752
## 2: 92           1st Qu.:2018-07-28 Charcoal Fabric :430
## 3: 140          Median :2018-07-30 Fire TV Stick   :340
## 4: 447          Mean    :2018-07-24 White Dot       :268
## 5:2246          3rd Qu.:2018-07-30 Black Plus      :261
##              Max.    :2018-07-31 Black Show      :259
##              (Other)           :761
## review_text
## Length:3071
## Class :character
## Mode  :character
##
##
##
##
```

As shown in Figure 1, most of the reviews are concentrated at the end of July, probably due to the fact that in 2018 Amazon set the so-called Prime days, i.e. a discount period lasting from 16 to 17 July. Moreover, Amazon usually sends an email notification to the customer a few weeks after the purchase, asking them to publish the evaluation of the purchased product.

The Amazon Echo models include 14 variations which differ in types and colors. Specifically, there are 6 models:

- *Echo* (2nd gen.) in Charcoal Fabric, Heather Gray Fabric, Oak Finish, Sandstone Fabric and Walnut Finish;
- *Echo Dot* (2rd gen.) in black and white;
- *Echo Plus* (1st gen.) in black and white;
- *Echo Show* (1st gen.) in black and white;

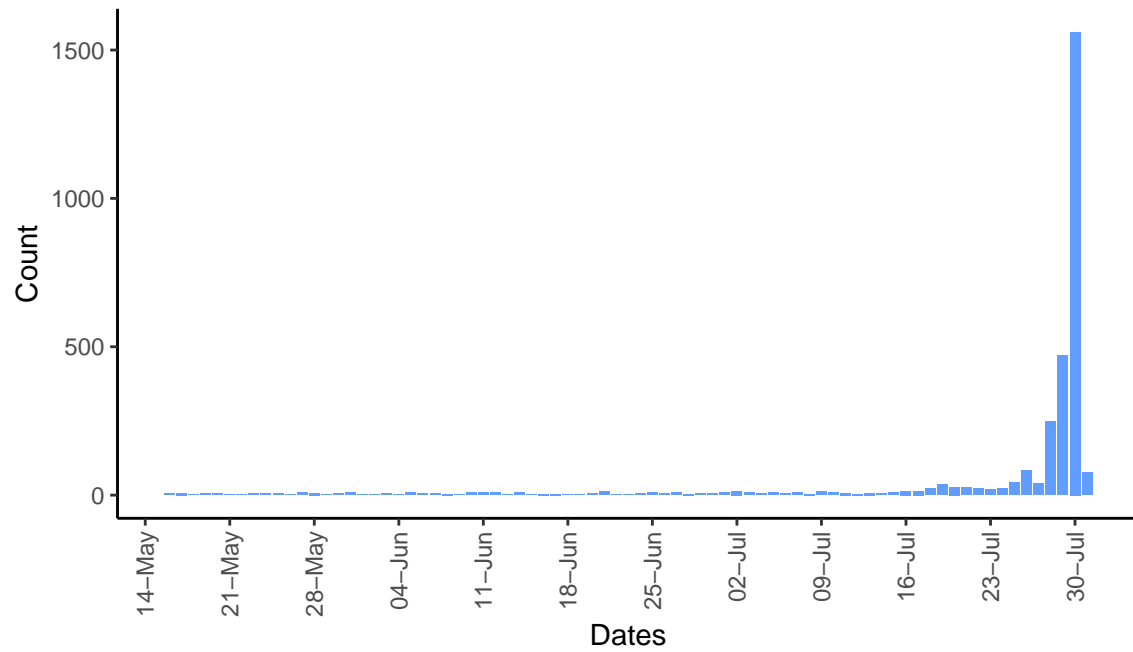


Figure 1: Number of Amazon Alexa reviews by date.

- *Echo Spot* in black and white, and finally
- *Fire TV stick*.

The first five have very similar functions, differing mainly in the presence or lack of a screen and in its size. The last product, on the other hand, allows the TV control via Alexa.

As shown in Figure 2, presumably the most purchased products were the Black Echo Dot, the Charcoal Fabric Echo and the Fire TV stick. It can be seen that these are also the products in the Alexa line that tend to have the lowest prices.

Moreover, the Figure @ref(fig:prod-rate-var) shows that most ratings are positive for each model in each of their variation.

It is though decided to isolate the model from the color by renaming the levels of the model variable in order to have a clearer overview not influenced by the price that varies between each model's color.

```
df <- df %>% mutate(product_model=recode(product_variation, "Charcoal Fabric"="Echo",
    "Heather Gray Fabric"="Echo", "Oak Finish"="Echo",
    "Sandstone Fabric"="Echo", "Walnut Finish"="Echo",
    "Black Dot"="Echo Dot", "White Dot"="Echo Dot",
    "Black Plus"="Echo Plus", "White Plus"="Echo Plus",
    "Black Show"="Echo Show", "White Show"="Echo Show",
    "Black Spot"="Echo Spot", "White Spot"="Echo Spot"))
```

In Figure 4 the percentage of Echo reviews is shown and curiously the models Echo Plus, Show and Spot, together with

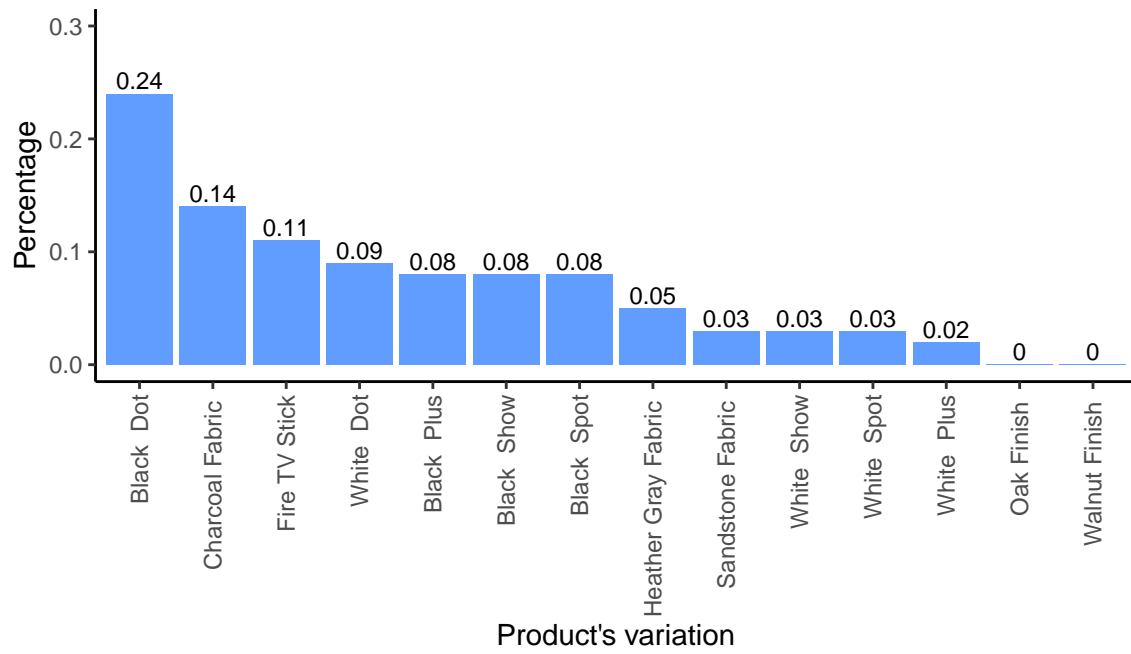


Figure 2: Percentage of Amazon Alexa reviews by product variation.

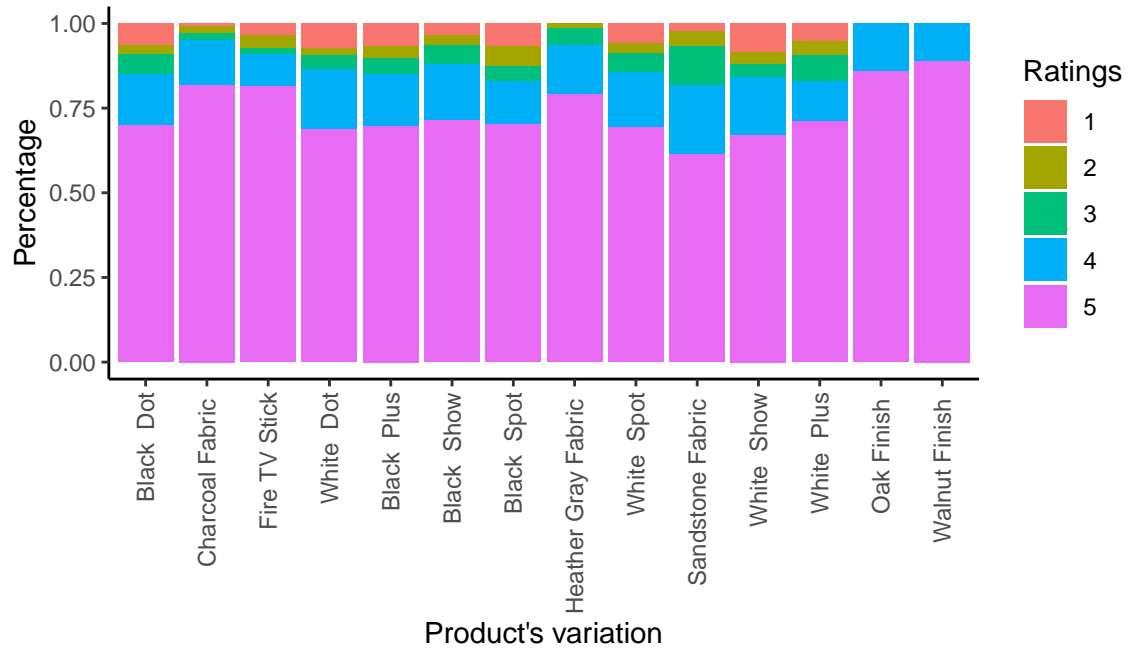


Figure 3: Ratings' percentage of Amazon Alexa reviews by product variation.

Fire TV stick have exactly the same number of reviews, which suggests that the source has included some condition in the data collection.

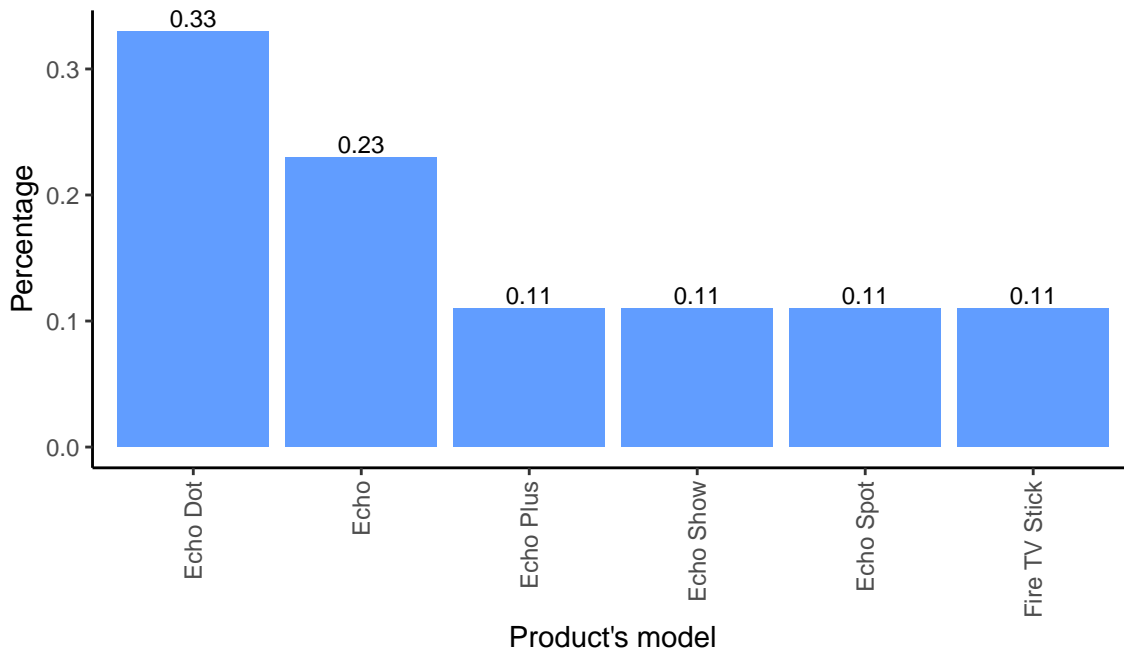


Figure 4: Percentage of Amazon Echo reviews by product model.

However, as seen in Figure 5 the aggregated data do not show any particular differences in terms of ratings: roughly speaking, the highest ratings are proportionally the same in all models, even if Echo Spot and Dot have more ratings below 2 than the other models. Echo and Fire TV Stick, instead, are the models with more positive ratings ( $>3$ ).

### 3.4 Opinion mining

Opinion mining, also called sentiment analysis, on reviews has been carried out in order to answer the research questions. This is a text analysis technique that uses natural language processing (NLP) to identify the sentiment behind a text. R provides many libraries to perform it, in this analysis **quantenda** v3.0 has been used along with **textstem** and **tidytext**.

In order to work with text data, the first step is to create a corpus from the original text. Quanteda's corpus, as reported in its documentation, is a "*library* of original documents converted to plain, UTF-8 encoded text and stored with document-level meta-data". The latters are named **docvars** and in this case are product's model and rating.

```
# Create corpus
reviews_corpus <- corpus(df$review_text,
  docvars=data.frame(product=df$product_model,
    rating=df$product_rating))
```

As shown in the summary above, each corpus' document is a review to which has been associated the corresponding model



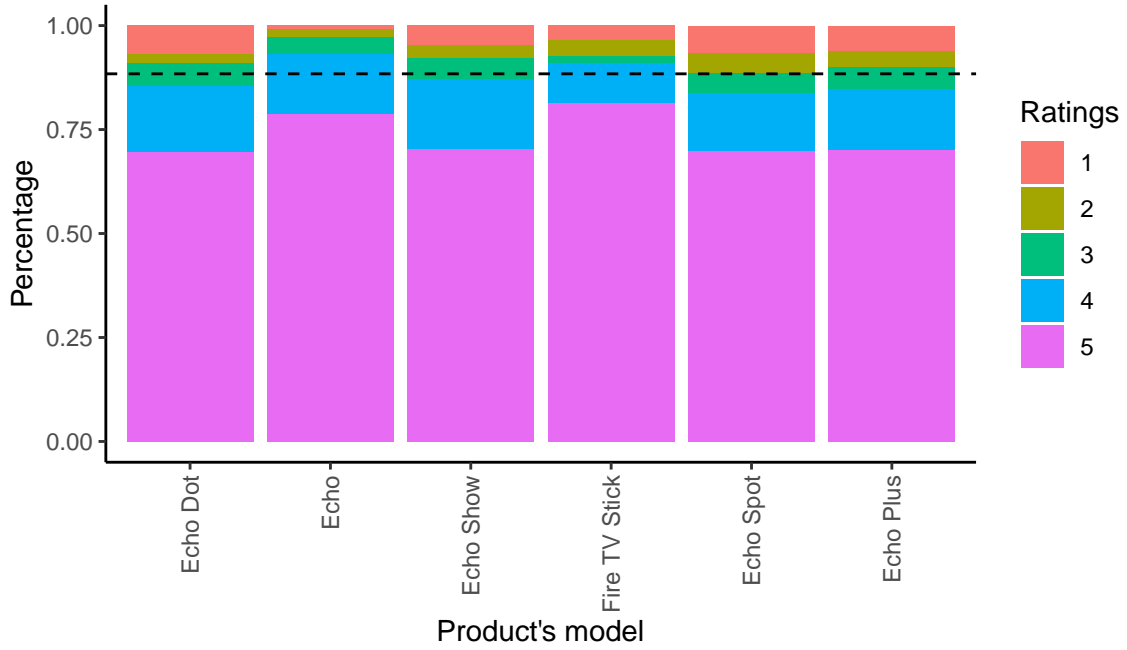


Figure 5: Ratings' percentage of Amazon Alexa reviews by product model.

and rating.

## Corpus consisting of 3071 documents, showing 5 documents:

##

##	Text	Types	Tokens	Sentences	product	rating
##	text1	4	4	1	Echo	5
##	text2	3	3	1	Echo	5
##	text3	35	41	2	Echo	4
##	text4	36	38	3	Echo	5
##	text5	1	1	1	Echo	5

The reviews are now ready to be tokenized to perform analyses. Tokens are the single characters contained in a sentence, including punctuation and symbols. It is worth noting that the summary above already contains the number of tokens for each document, but they need to be cleaned. Indeed, it was created a function which:

- set conditions for subsetting depending on docvars;
- removes punctuation, numbers, symbols and separators;
- removes the patterns “#34” and “.” since **quanteda** was not deleting them;
- forces the words “alarm clock” to be together as they were treated as distinct words changing their meaning;
- removes stopwords using **tidytext** list, as **quanteda** also removed negations. Moreover, it was also decided to remove products' names since they do not add any specific useful information;

- lemmatizes the words using `textstem` list, as `quanteda` only allows to stem words, operation which tends to lose the text's context and so not applied in this analysis.

```
my_toks <- function(my_corpus, subset_prod=FALSE, sub_p="",
                    subset_rat=FALSE, sub_r=""){

  # Subset conditions
  ifelse(subset_prod, my_corpus <- my_corpus %>% corpus_subset(product %in% sub_p), my_corpus)
  ifelse(subset_rat, my_corpus <- my_corpus %>% corpus_subset(rating %in% sub_r), my_corpus)

  # Tokens creation
  ret <- my_corpus %>%
    quanteda::tokens(remove_punct=TRUE, remove_numbers=TRUE, # quanteda's tokenizer
                     remove_symbols=TRUE, remove_separators=TRUE) %>%
    tokens_remove(pattern=c("#34", ".")) %>% # remove some pattern
    tokens_tolower() %>% # lower case conversion
    tokens_compound(pattern=phrase("alarm clock")) %>%
    tokens_remove(pattern=c(tidytext::stop_words$word, # remove stopwords
                           "echo", "dot", "plus", "show", "spot", # remove other words
                           "fire", "tv", "stick", "firestick",
                           "amazon", "alexa")) %>%
    tokens_replace(pattern=lexicon::hash_lemmas$token, # lemmatization
                   replacement=lexicon::hash_lemmas$lemma)

  return(ret)}

```

The function is then applied to reviews' corpus as follows:

```
reviews_toks <- my_toks(reviews_corpus)
```

Tokens can now be converted in document-feature matrix, pivotal format for conducting analysis.

```
reviews_dfm <- reviews_toks %>% dfm()
reviews_dfm
```

```
## Document-feature matrix of: 3,071 documents, 2,922 features (99.76% sparse) and 2 docvars.
```

```
##           features
```

```
## docs    love game answer question correctly wrong home lot fun yr
```

```
## text1    1    0    0    0    0    0    0    0    0    0
```

```
## text2    1    0    0    0    0    0    0    0    0    0
```



used to watch movies. And, finally, Echo seems to be used mainly to listen to music.

Moreover, we can visualize the network of co-occurrences between words using the following function, which allows us to see the connection between the reviews' top 30 features. The words' font size changes depending on the frequency of words: the biggest appear more frequently.

```
network <- function(toks){  
  fcmat <- fcm(toks, context="document", tri=FALSE)  
  feat <- names(topfeatures(fcmat, 30))  
  fcm_30 <- fcm_select(fcmat, pattern=feat)  
  textplot_network(fcm_30, min_freq=0.8, vertex_labelsize=2*(rowSums(fcm_30)/min(rowSums(fcm_30))))  
}
```

The Figure 7 confirms the great usage of the smart speakers to interconnect devices into the house, listening to music and set alarms. In this case the separation between models is not underlined, but it would be easy to state that the biggest terms refer to models for which there are more reviews.

```
set.seed(101)  
network(reviews_toks)
```

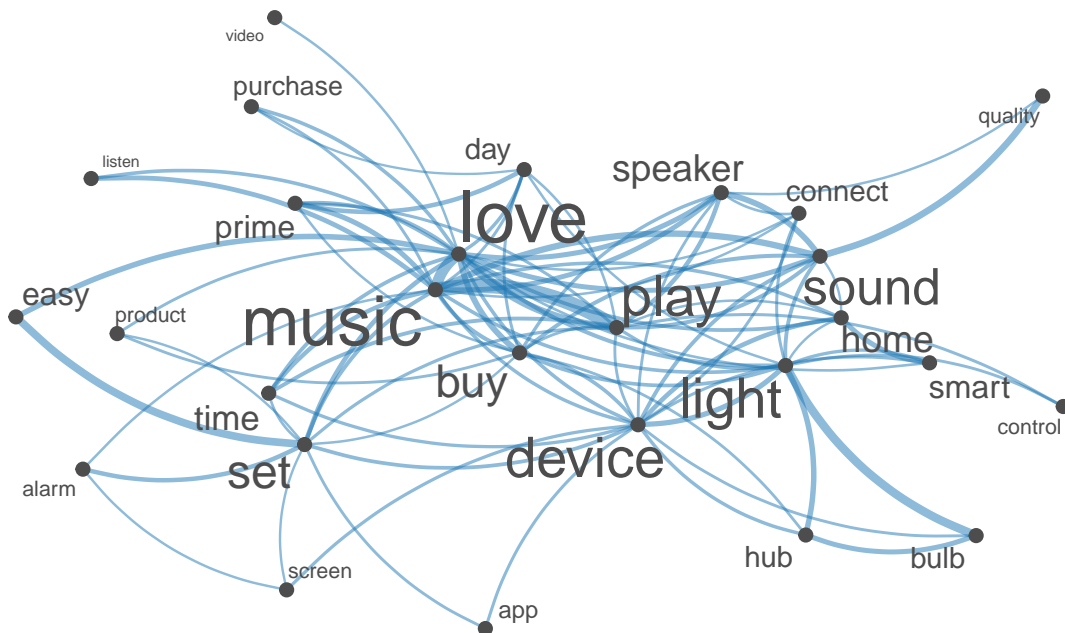


Figure 7: Network of co-occurrences between reviews' words.

As mentioned above, in order to answer to the second research question, are customers concerned about privacy, opinion mining has been performed using NRC Word-Emotion Association Lexicon. The latter includes a list of English words and their associations with eight emotions: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust; and two

sentiments: negative and positive.

To start, the dfm containing the number of words contained in each document is transformed into a dataframe, then it is compute the total number of time each word appear and transposed the dataframe. At this point the NRC lexicon has been applied to classify each word. Finally, a filter is applied to keep only sentiments and emotions useful to answer to the second research question.

```
words_n_sent <- reviews_dfm %>%
  convert(to="data.frame") %>%           # converting dfm into dataframe
  summarise(across(where(is.numeric), ~ sum(.x))) %>% # sum across columns
  t() %>%                                # transposition
  as.data.frame() %>%
  rename("n"="V1") %>%
  rownames_to_column(var="word") %>%     # count and sorts words
  inner_join(get_sentiments("nrc"), by="word") %>%
  filter(sentiment %in% c("positive", "negative", "trust", "fear"))

head(words_n_sent)
```

```
##      word    n sentiment
## 1    love 1060  positive
## 2 question   82  positive
## 3   wrong   15  negative
## 4    fun   152  positive
## 5  music   538  positive
## 6   gift    52  positive
```

As emerges in Figure 8 and already highlighted in exploration analysis, most of the reviews contain positive words and the 19% are trusting words. Although the analysis shows that just the 6% of words are associated to fear.

```
sent_count_df <- words_n_sent %>%
  group_by(sentiment) %>%
  summarise(n=sum(n)) %>%
  mutate(perc =round(n/sum(words_n_sent$n), 2))

ggplot(sent_count_df, aes(x=sentiment, y=n)) +
  geom_bar(stat="identity", fill="#619DFF") +
```

```
geom_text(aes(label=perc), size=3, position=position_dodge(width=0.9), vjust=-0.25) +
theme_classic() +
theme(legend.position="none") +
labs(x="Sentiments",
      y="Scores")
```

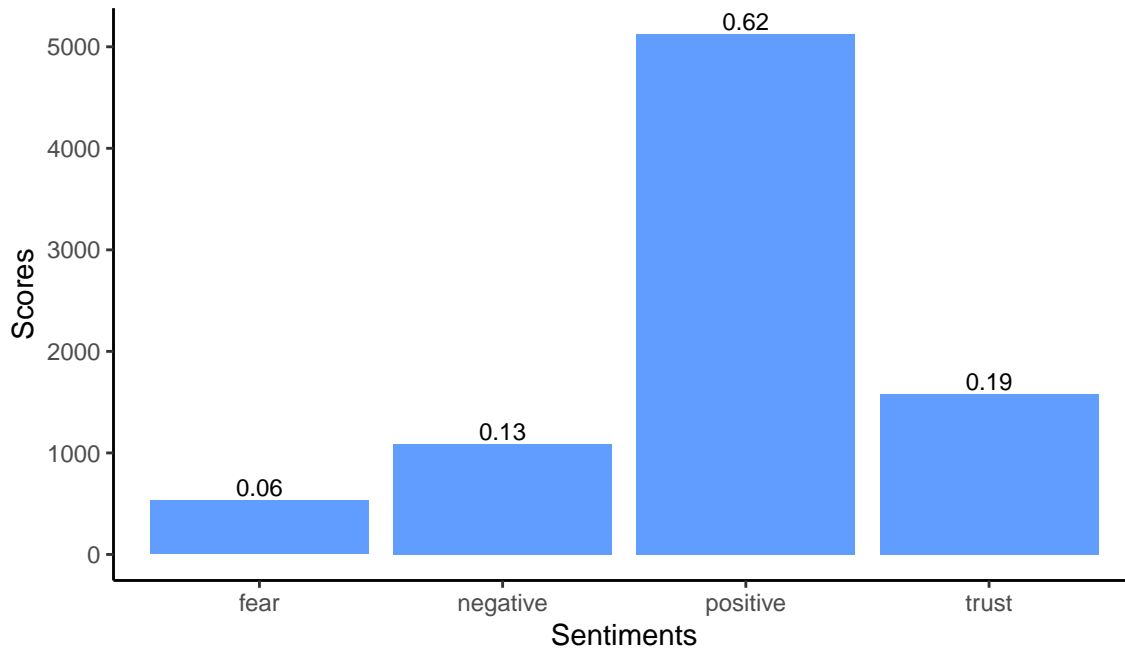


Figure 8: Sentiment of Alexa's products reviewers.

In Figure 9 are shown the first 10 words associated with the selected sentiments and emotions.

```
words_n_sent %>%
  group_by(sentiment) %>%
  slice_max(n, n=10) %>%
  mutate(word=reorder(word, n)) %>%
  mutate(sentiment=factor(sentiment, levels=c("positive", "trust", # ordering output
                                              "negative", "fear"))) %>%

  ggplot(aes(n, word, fill=sentiment)) +
  geom_col(show.legend=FALSE) +
  facet_wrap(~sentiment, scales="free_y") +
  labs(x="Contribution to sentiment and to emotion",
       y=NULL)
```

Although, some of words associated with fear are probably misclassified. For instance “alarm” is probably referred to

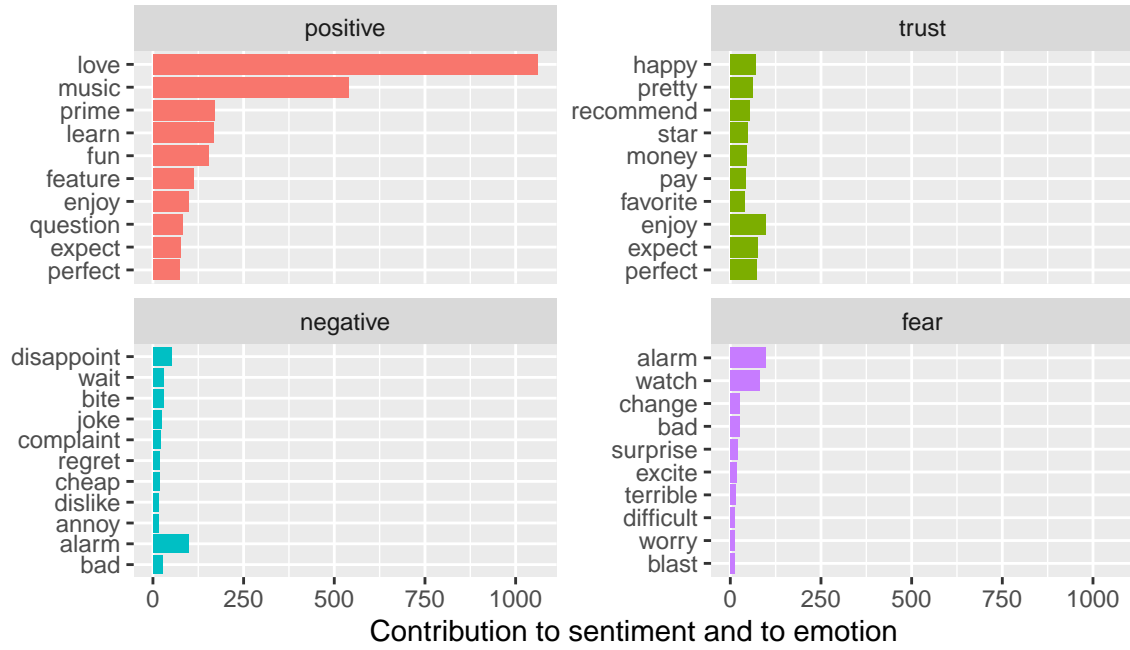


Figure 9: Sentiment associations with Alexa’s products reviewers’ words.

“alarm clock” and does not denote a level of concern. The only words worth of noting is “worry” but just with this analysis it is difficult to understand its context. The other words are mainly associated to product criticism, probably left by the small number of unsatisfied customers.

### 3.5 Results

The analysis shows that each product in the Echo line takes on a different role in the home: from a smart control hub to a simple clock, acting as a true digital assistant. Furthermore, the analysis of the reviews does not show that there are any privacy concerns, but there is probably not enough data to prove this.

## 4 Conclusion

The tools used to conduct the analysis are very powerful and apply to different contexts and situations. The wordcloud only gives an idea of how the corpus considered is composed and, in this case, it was useful to identify the uses made of the products. Regarding the classification of general sentiment, some aspects of the analysis could be improved. In fact, some misclassifications are noticed and the synthesis that emerges from the reviews seems to be inaccurate. Sentiment analysis tools at these levels are still being developed and become more accurate as the years go by. For future projects it would be interesting to base similar analyses on a larger dataset, in order to be able to distinguish the most appreciated features from the most criticised ones. In addition, for a more accurate answer to the second research question, it would be necessary to base the analysis on other sources: it is unlikely that people who have a real concern about privacy are

Amazon Echo customers or, if they were, have left a review on the website whenever they realized it.

#### Sources

- USA Today Tech, URL [usatoday.com](https://www.usatoday.com)
- H. Chung, M. Iorga, J. Voas and S. Lee, “Alexa, Can I Trust You?” in *Computer*, vol. 50, no. 9, pp. 100-104, 2017.
- European Commission, “The rise of Virtual Personal Assistants Digital”, *Transformation Monitor*, 2018.