MKTG 788 Exam 2

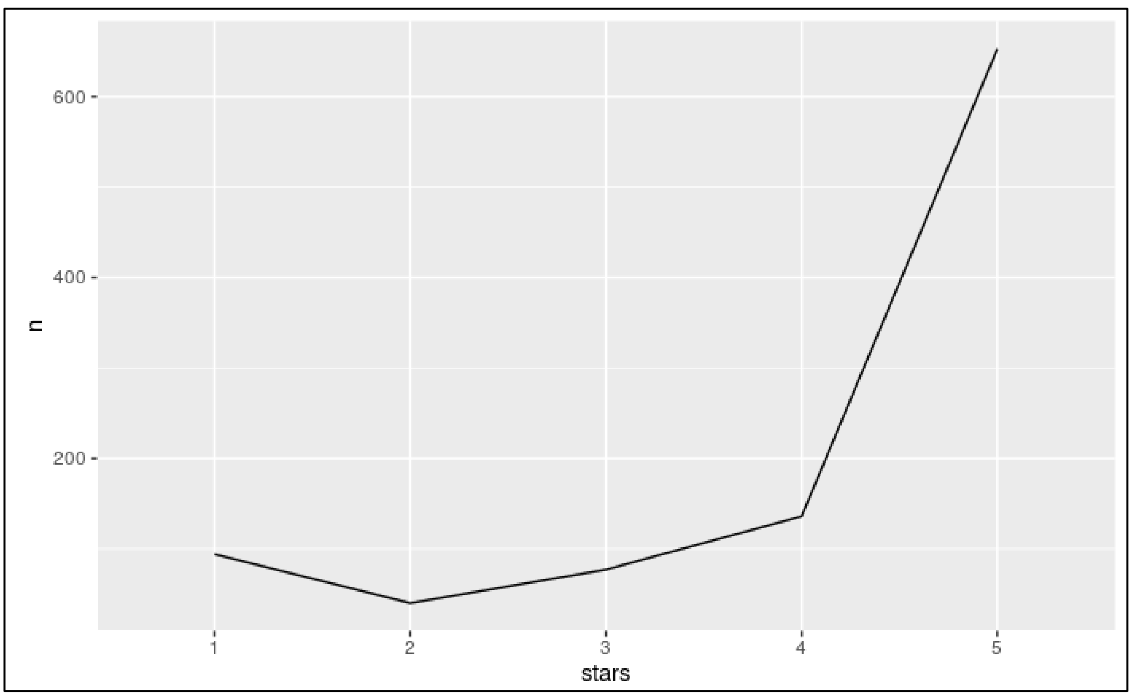
Hutzler Pepper Saver, Red

Text Analysis of Amazon Reviews

The dataset used for this analysis is B005BPZCUC_Reviews for Exam 2.csv. The dataset consists of 1000 reviews with 9 variables. Here is a look at the data:

```
reviews_orig          1000 obs. of 9 variables
  ID : chr "7" "10" "15" "18" ...
  prod : chr "Hutzler Pepper Saver, Red" "Hutzler Pepper Saver, Red" "Hutzler Pepper S
  title : chr "Great Concept; Not So Easy To Use" "Love this onion saver!" "and just I
  author : chr "Cynthia June" "Mariposa" "Christopher Latson" "JAB" ...
  date : chr "8-Jan-19" "5-Apr-16" "17-Mar-17" "19-Jul-16" ...
  review_format: chr "Color: Lime SaverVerified Purchase" "Color: Red Onion SaverVeri
  stars : chr "4" "5" "5" "5" ...
  comments : chr " The containers that look like their contents is an awesome idea, I
  n_helpful : chr "1" "1" "1" "1" ...
```

The following plot show the distribution of reviews as per their star ratings:



Around 650 reviews have a 5-star rating followed by ~130 reviews with a 4-star rating. Reviews with 1,2&3-star ratings together are ~200 in number.

We will begin by grouping the reviews by their Star rating (1-5) and then analyzing their sentence level sentiments using the Tyler Rinker's dictionary. Tyler Rinker is the author of sentimentr-the package used to generate the sentence level sentiment score.

Below is a box plot for the review sentiment scores grouped by their star ratings. In the plot below, we can see that the medians of review sentiment scores are well above or below the Neutral sentiment (0.0). This aligns with the star ratings. Few outliers can be seen in the sentiment scores. This suggests high use of positive or negative sentiment words in that review. Surprisingly, some outliers in the 4&5-star rating groups lie in the negative side of the plot.

Medians of average sentiments of 1- & 2-star rated reviews lie below the neutral sentiment line, while median of 3-star rated review sentiments lies on the neutral line. Other 4 & 5 star rated reviews lie well above the neutral sentiment scores.

We can thus say that the sentiment scores and the star ratings of the reviews go well with each other. There are few contradictions (outliers) which, however, do not affect the results much.

The minimum (the smallest sentiment in the data set) is shown at the far bottom of the chart, at the end of the bottom "whisker."

The median is shown as a line somewhere in the middle of the box.

First quartile, Q1, is the bottom part of the box .

Third quartile, Q3, shown at the top of the median.

The maximum (the largest sentiment in the data set), shown at the top of the box.

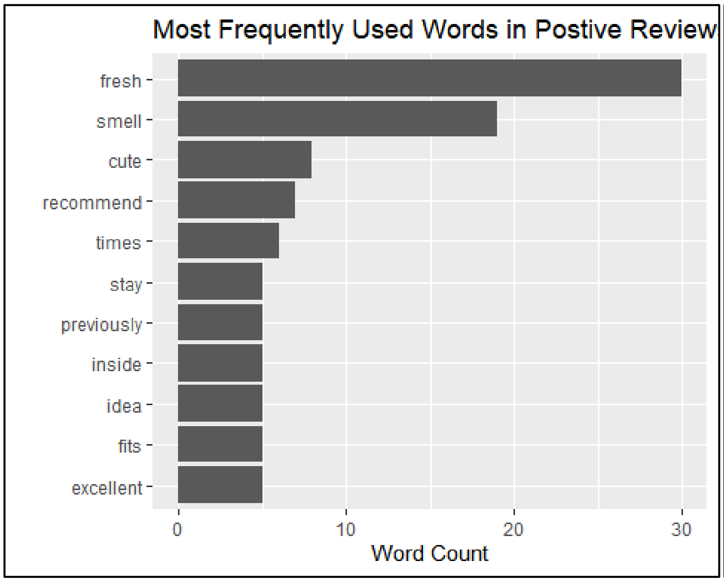
We will now divide the reviews as positive and negative based on their star ratings. Since sentiment scores of reviews with 1,2&3-star ratings have at least one quadrant below or intersected by the neutral sentiment line, we will consider them as negative review. Sentiment scores of reviews with 4&5-star ratings lie well above the neutral sentiment line; hence we will consider them as positive reviews. We get 789 positive reviews and 211 negative reviews as shown below.

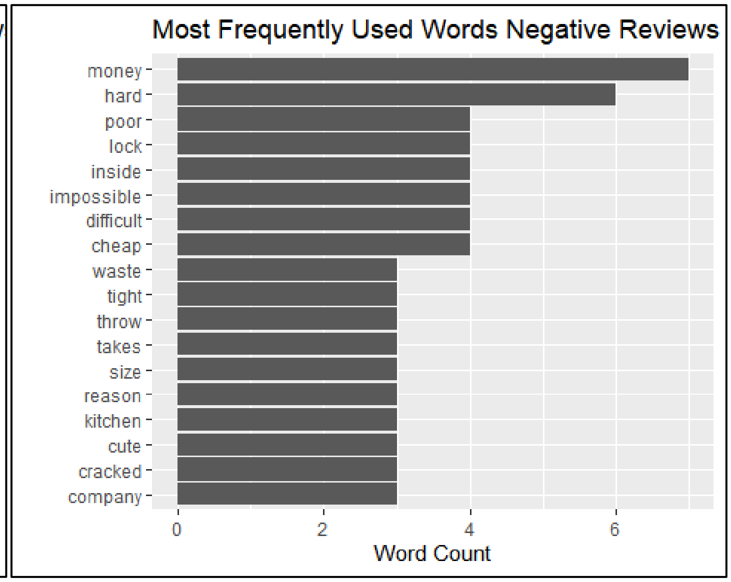
```
Rows: 789
Columns: 3
$ stars    <chr> "4", "5", "5", "5", "5", "5", "5", "5", "5", "5", "5", "4",…
$ comments <chr> " The containers that look like their contents is an awesom…
$ ID       <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, …
Rows: 211
Columns: 3
$ stars    <chr> "1", "1", "3", "1", "1", "1", "1", "1", "1", "1", "1", "1",…
$ comments <chr> "The design to open this saver is ridiculous.It takes a lot…
$ ID       <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, …
```

Next, we will remove some undesirable words from both the datasets that might cause bias in the analysis.

"saver", "onion", "tomato", "lemon", "fridge", "garlic", "onions", "container", "pepper", "plastic", "half", "lemons", "product", "time", "close", "time", "store", "bought", "tomatoes", "grapefruit", "refrigerator", "days", "easily", "easy", "item", "peppers", "bottom", "storage", "savers", "perfect", "containers", "lime", "limes", "savers", "keeper", "pieces", "food", "makes", "save", "keeping", "love", "nice", "holder", "week", "save", "hold", "bags", "produce", "clean", "helps", "vegetable", "fruit", "wrap", "veggies", "holds"

Here is a plot that shows the most frequent words used in both the datasets:





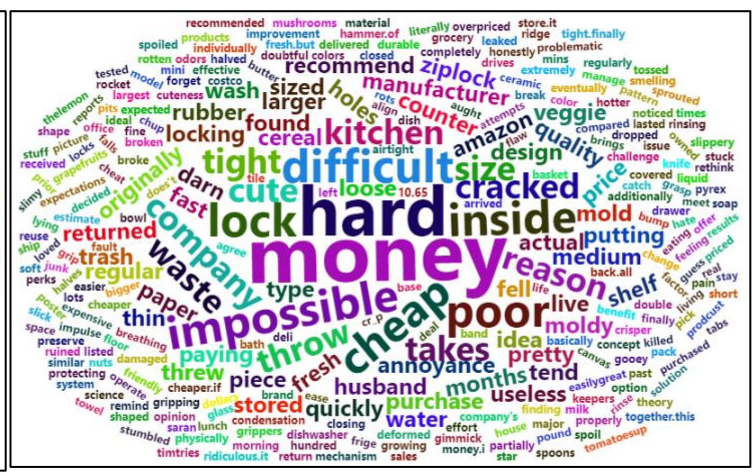
In the positive reviews, the reviewers have mostly written about the freshness, smell/odor, and 'cute' appearance of the pepper saver. Some have also mentioned about recommending the pepper saver

In the negative reviews, reviewers have written about money which implies that there is possibility of not finding the product "value for money". There are less chances of customers finding the pepper saver a "value for money" product in the negative reviews. Frequent use of words like "hard", "difficult" and "impossible" suggest on the usability of the product.

Word clouds

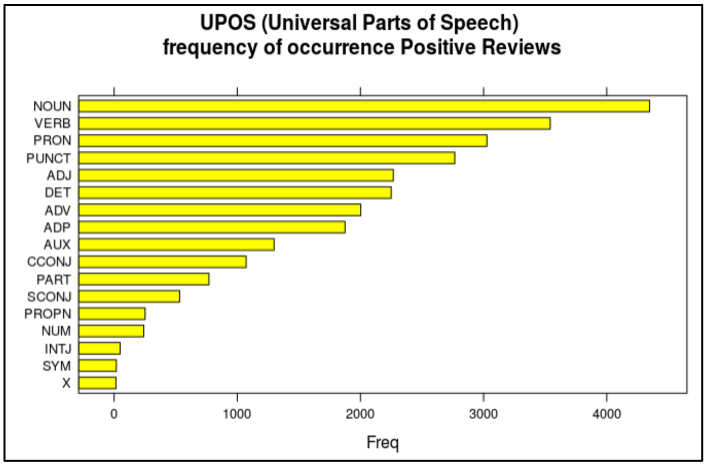
Positive Reviews

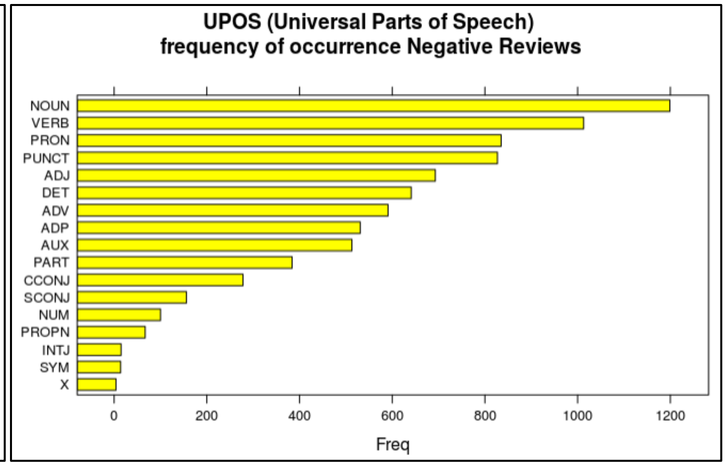
Negative Reviews



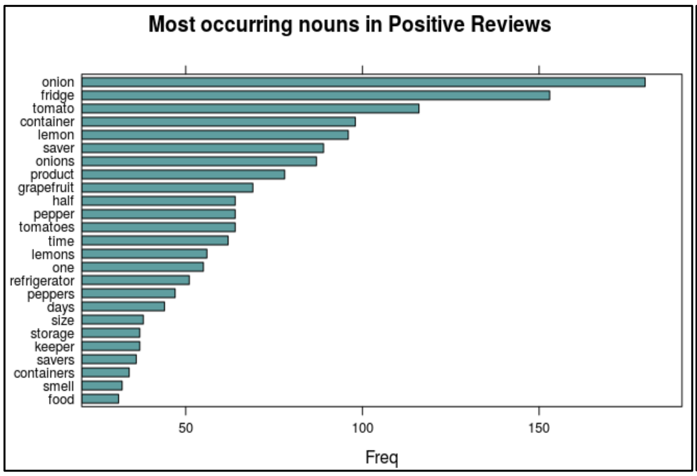
The word clouds above display some prominent words in both the datasets. Topics like “smell”, “fresh” are most frequent in the positive reviews, while topics like, “money”, “hard”, “impossible”, “cheap”, “poor” are some of the frequent topics among the negative reviews.

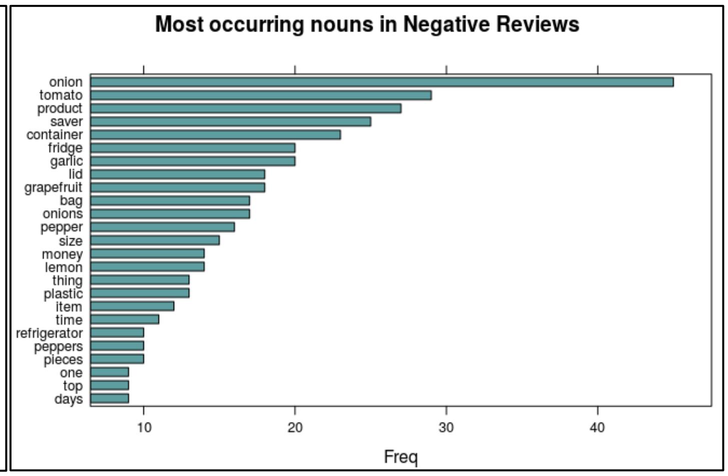
Speech analysis using UDPIPE and RAKE



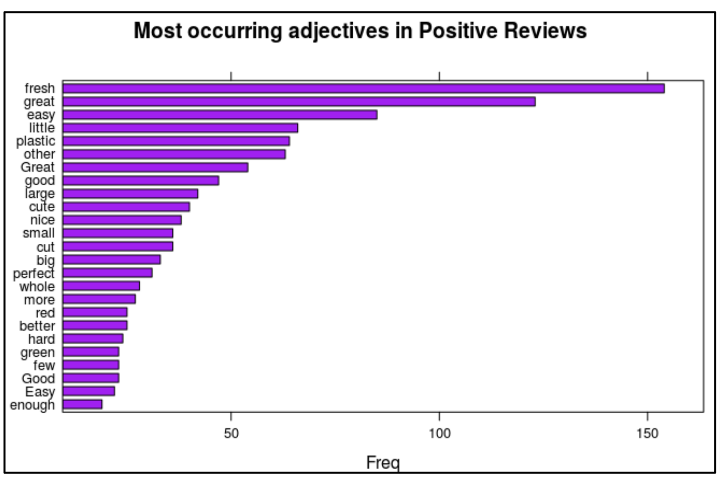


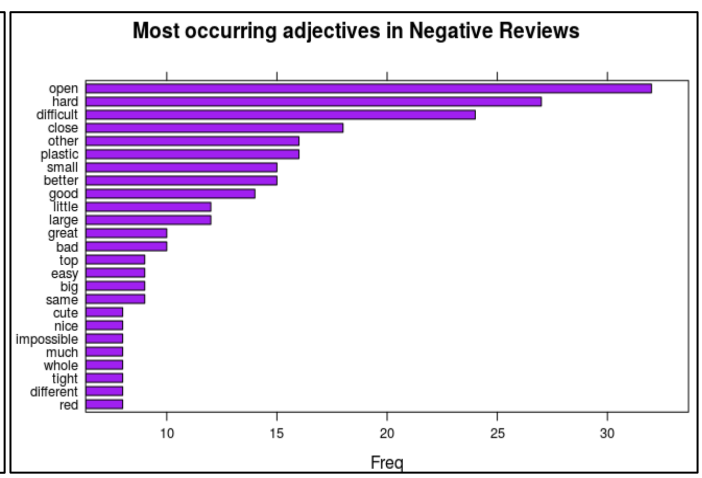
The plots above display the frequency of different parts of speech like nouns, verbs etc. The order of different parts of speech is almost similar in both the positive and negative reviews. During the analysis, we must consider the magnitude of reviews in each category, i.e. 789 positive reviews and 211 negative reviews in both datasets.



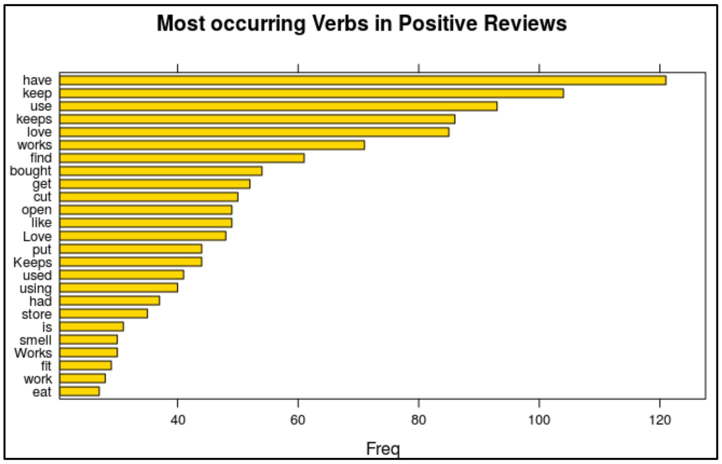


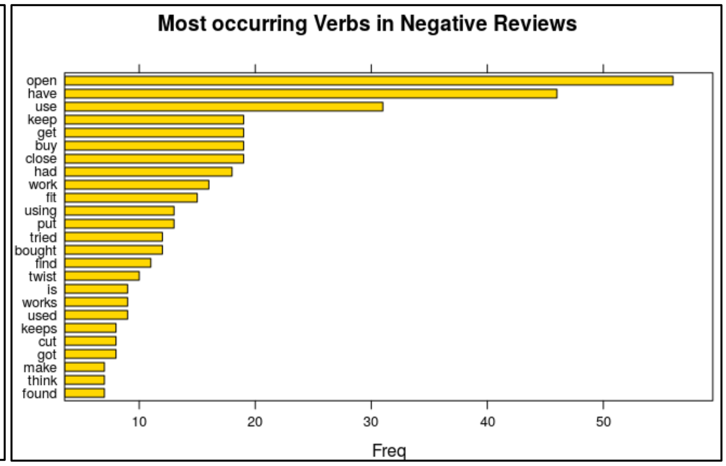
"Onion" and "tomato" are some of the most frequently used nouns in both the reviews. Surprisingly "pepper" is not amongst the top.



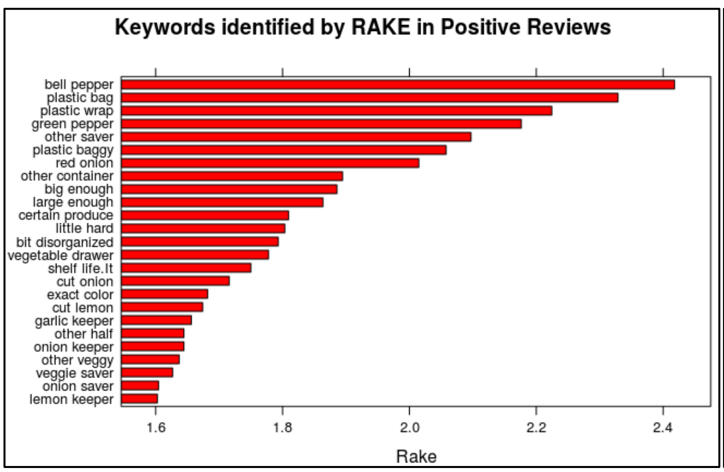


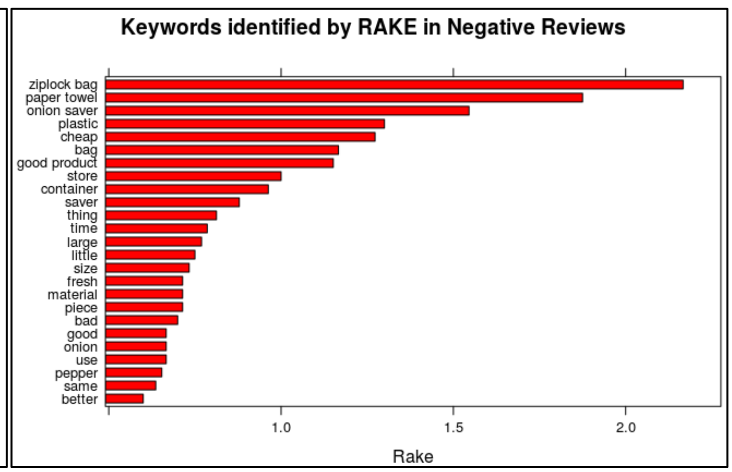
Positive reviews mostly have frequent words that talk about freshness, ease, size, and "cute" appearance of the pepper saver, while the negative reviews mostly contain frequent words that express "difficulty" in opening and closing the pepper saver. This contradicts the positive reviews that mention ease. However, after considering the frequency of these words in each category, we can say that overall, the pepper saver is easy to use.



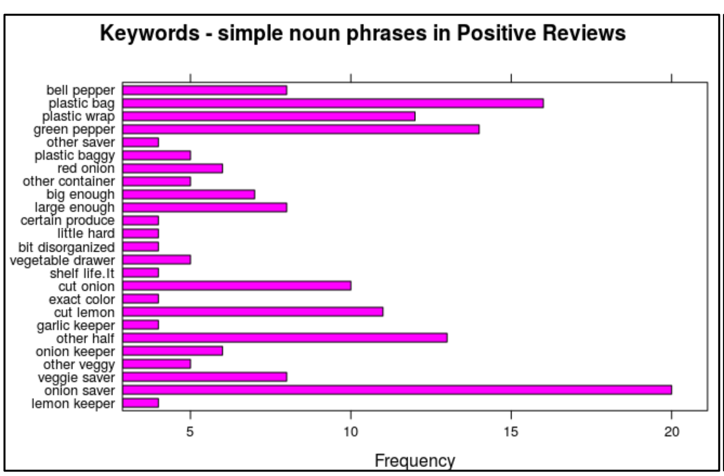


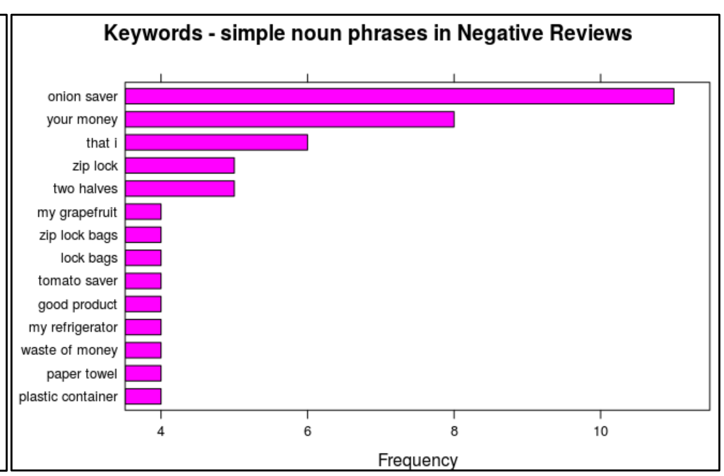
Some of the frequent verbs in the positive reviews are “love”, “open”, “smell”. These suggest that most of the user’s “love” the pepper saver. They might also find it easy to “open” in terms of usability. The veggies must “smell” good when stored in the pepper saver. While the most frequent verb in the negative reviews is “open”. Also, the verb “close” is frequent. These verbs suggest that a set of users find the usability of the product, i.e. opening and closing of pepper saver difficult.



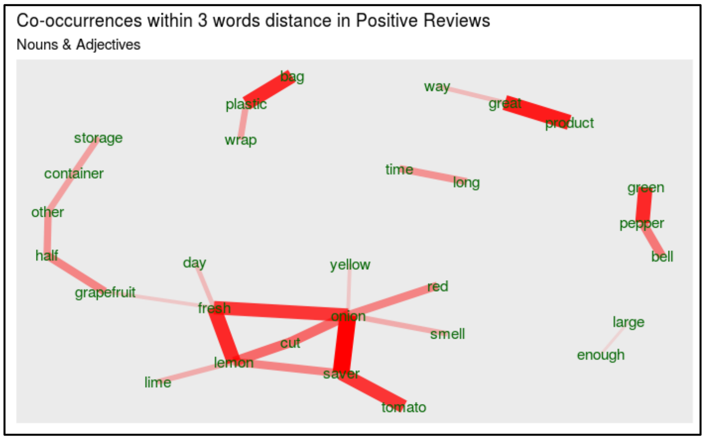


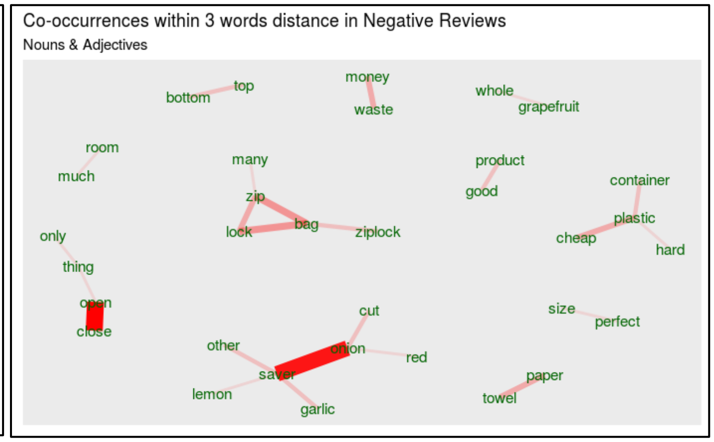
Some of the keywords in positive reviews are “big enough”, “large enough” suggest that the reviewers find the pepper saver adequate for storage. One thing tp notice is that the dataset is contaminated and might have reviews related to other veggie savers.





The simple noun phrases in negative reviews mostly mention other storage products like “zip lock bags”. This might suggest that they find the alternatives better than the pepper saver.





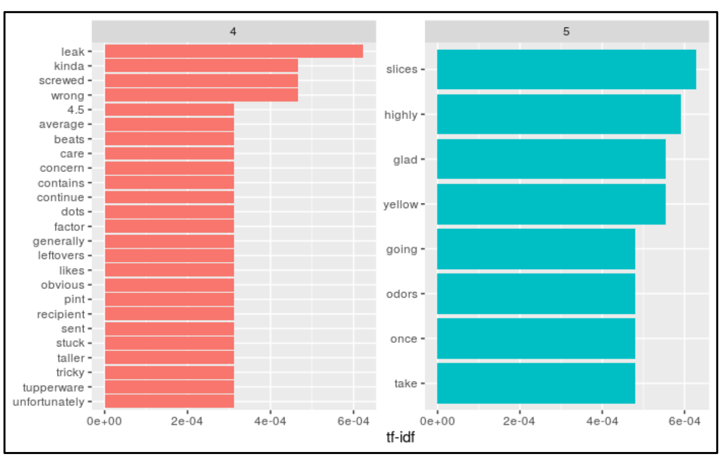
The 3-word distant co-occurring words in the positive reviews mainly suggest that the saver keeps the veggies fresh for long time. While in the negative reviews, phrases like money-waste, cheap-plastic-container, hard-plastic-container, open-close, top-bottom, zip-lock-bag bring forward some of the issues related to the saver. These 2- or 3-word distant phrases are easy to draw meanings from. For example, money-waste suggests that reviewers find that purchasing the saver is waste of money.

Topic Modeling

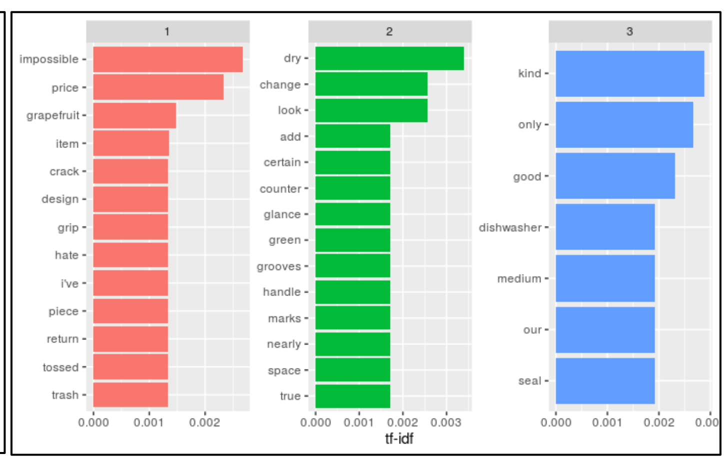
tf–idf or TFIDF, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus

First, we get the count of each word in each review.
Second, get the number of words per text input.
Third, combine these dfs and get the tf_idf & order the words by degree of relevance
Finally, we plot the 10 most informative terms per topic grouped by the grouping covariate

Positive Reviews



Negative Reviews



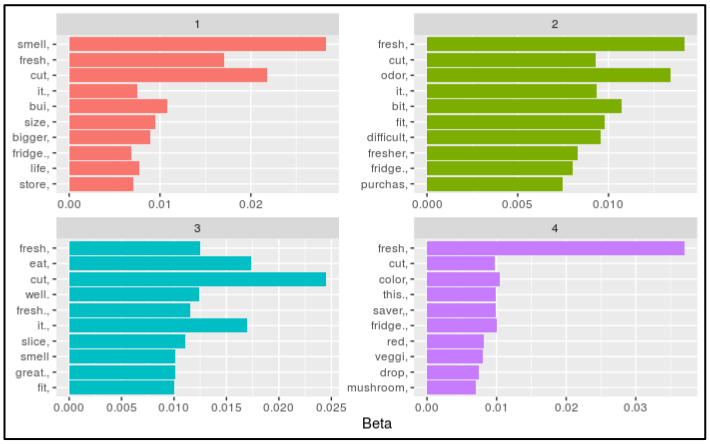
We have used star rating as the grouping covariate. Surprisingly the 4 star rated reviews have some topics that mentions “concern”, “screwed”, “leak” and “wrong” which seem negative. “Tupperware” is a competitor company that also produces such savers. This suggests some topics might contain comparison of the product with a competing company.

The tf-idf in negative reviews however justify the star ratings. Words like “price”, “crack”, “return”, “trash” suggest dissatisfied topics. The 3 star rated reviews seem to have neutral topics.

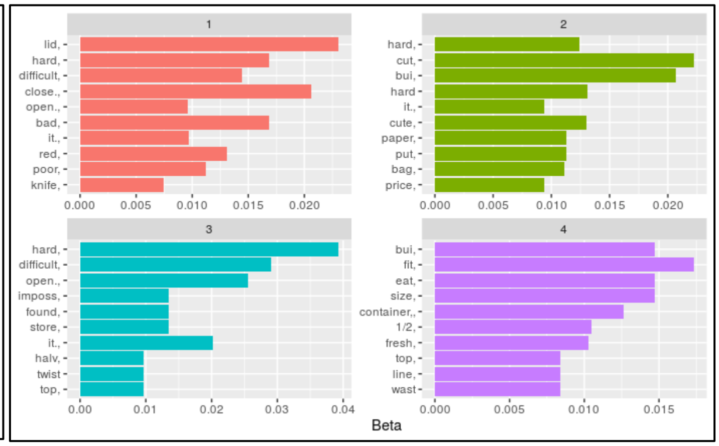
LDA

Latent Dirichlet allocation (LDA) is a particularly popular method for fitting a topic model. It treats each document as a mixture of topics, and each topic as a mixture of words. This allows documents to "overlap" each other in terms of content, rather than being separated into discrete groups, in a way that mirrors typical use of natural language.

Positive Reviews



Negative Reviews



This visualization lets us understand the four topics that were extracted from each of the positive and negative reviews.

- The most common words in topic 1 of positive reviews are "smell", "fresh", "size", "bigger".
- All the topics show some overlap and mostly relate to "smell", "fresh", "cut" or "fridge".

- The most common words in topic 1 of negative reviews are "hard", "difficult", "close", "open" which again suggest on the usability of the saver.
- Topic 2 has words like "cute", "price" and some overlap from topic 1 words. Topic 2 suggests that a user might find the product "cute" in appearance, but the word "price" suggests that it is not value for money.

Topics in negative reviews have less overlap of words as compared to positive reviews.

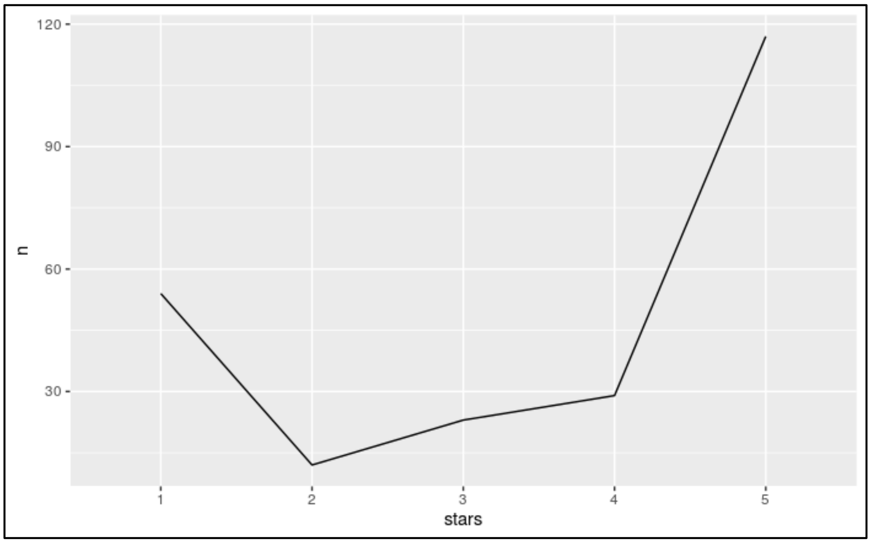
Helpful Reviews

The dataset also includes a column called 'n_helpful'. This column has a count of upvotes for a review. The values range from 1 to 36 in our dataset. We can do some analysis on the reviews that have at least 1 upvote as "Helpful". This will further help us to understand what features of a review make it helpful or not helpful for other potential customers.

For this let us divide the dataset into two, one marked as helpful and other that is not marked as helpful. There are 235 reviews out of 1000 marked as 'helpful' at least once.

```
Rows: 235
Columns: 4
$ stars     <chr> "4", "5", "5", "5", "5", "5", "5", "5", "5", "5", "1", "1", "3", "1", "5", "1", "1", "1", "…
$ comments  <chr> " The containers that look like their contents is an awesome idea, I couldn't wait to try o…
$ n_helpful <chr> "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "1", "…
$ ID        <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, …
Rows: 765
Columns: 4
$ stars     <chr> "5", "5", "1", "5", "1", "3", "5", "4", "5", "5", "5", "4", "4", "5", "5", "5", "5", "5", "…
$ comments  <chr> "This is a very cute container to put in the frig and to store my mushroom. My only complai…
$ n_helpful <chr> "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "0", "…
$ ID        <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, …
```

Here is the distribution of star ratings for the helpful reviews:



Again, most of the reviews that were found helpful were rated as 5-star followed by 1-star.

Comparison of Parts of speech and topic models on both helpful and not_helpful datasets give similar results.

However, one striking difference between both is the length of the review or the number of words used in a review.

To find this we add a new column in these datasets that consists of the total number words used in these reviews. Please note that these reviews are raw and unfiltered.

Now let us find the mean of the total number of words used in each of these reviews.

```
```{r}
mean(nothelpful_reviews$total)
mean(helpful_reviews$total)
```

 [1] 20.78431
 [1] 62.58298
```

It appears that on an average, the length of ‘helpful’ reviews is ~63 words while the ‘not_helpful’ reviews have an average length of just ~21 words. Thus, we can say that, the more descriptive and lengthier a review is, the more helpful it is found by the readers. In other words, long reviews, or reviews of ~63 words in length are considered to be more helpful.