Gelato/Sorbet Sentiment Analysis

Erin Alcott

2023-07-20

For this project, I wanted to identify what people like most about Talenti’s most popular products. To do so, I used lexicon-based sentiment analysis to classify words in product reviews as “positive” or “negative” (or neutral). I ranked Talenti products by their ratings and by the percentage of positive sentiment in their reviews. I then created word clouds to visualize the most commonly used words in positive reviews (3 stars or higher) and the most commonly used positive words in these reviews. Doing so helps give attention to relevant non-sentimental descriptors (e.g.: “chocolate,” “pint”) and relevant sentimental descriptors (e.g.: “smooth”).

Find my data source here: <https://www.kaggle.com/datasets/tysonpo/ice-cream-dataset>

# **CONCLUSIONS**

Talenti has mostly 5-star reviews and products with a median rating of 4.4, and only one product out of the 45 in the dataset falls below 3 stars. Talenti’s products are generally well-liked among their customers leaving reviews in the dataset. The top flavor in terms of high ratings and high sentiment percentage is the Organic Oak-Aged Vanilla Gelato, with a rating of 4.8 and 82% positive sentiment percentage across all reviews. Some of Talenti’s most popular products also include the Coconut Almond Chocolate Gelato, the Caramel Apple Pie Gelato, the Belgian Chocolate Gelato, the Organic Ginger Matcha Gelato, the Peanut Butter Fudge Sorbetto, and the Alphonso Mango Sorbetto.

Based on my analysis, Talenti’s top products are loved for their unique flavors and flavor combinations, their smooth and creamy texture, and their balanced qualities, including flavor combinations (i.e.: none overpower or are lacking), ingredient distributions, and the overall quality of the product. Talenti’s sorbettos offer dairy-free alternatives that reviews suggest may have similar textures to other dairy-containing frozen desserts. It’s likely these attract a dairy-free market, since many positive reviews referenced the desserts being “dairy-free” and/or “vegan.” Many reviews referenced flavors being “favorites,” which may mean unique and successful products are helping Talenti sustain reliable sales with groups that love specific products in particular. Market research pertaining to these topics may be insightful and interesting to assess.

Complaints for the lowest rated desserts typically involve poor quality upon consumption or a feeling of false advertising, often as a result of imbalanced qualities/flavor combinations of the dessert. Potential solutions could include increased customer service interactions or improved quality sustenance instructions on packaging/shipment or the containers themselves. At the worst case, recipe changes could be made to improve these qualities of the product where possible. More research (e.g.: market research, quality control research) would need to be done before creating definitive solutions.

Of interest to me was that many of the Top 10 flavors by Rating included chocolate, and many lower-rated flavors in this regard were pie/crumble/cookie-based flavors. A future project may include correlation testing between chocolate ingredients and high ratings, as well as wheat flour and low ratings.

# **Data Preparation**

suppressWarnings({  
 rm(list=ls())  
 library(textdata)  
 library(tidyverse)  
 library(tidytext)  
 library(textclean)  
 setwd("C:\\Users\\eaalc\\OneDrive - Umich\\Ice Cream")  
   
 ice\_cream <- read.csv("archive (10)\\talenti\\products.csv")  
 review <- read.csv("archive (10)\\talenti\\reviews.csv")  
 ice\_cream <- data.frame(ice\_cream)  
 review <- data.frame(review)  
})

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --  
## v dplyr 1.1.2 v readr 2.1.4  
## v forcats 1.0.0 v stringr 1.5.0  
## v ggplot2 3.4.2 v tibble 3.2.1  
## v lubridate 1.9.2 v tidyr 1.3.0  
## v purrr 1.0.1   
## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

# **Data Exploration**

First, I looked at the first five rows of the two datasets. Then I looked at the flavors of ice cream in the dataset. There are 45 total ice cream flavors. Finally, because the review dataset uses flavor “keys” rather than flavor names, I did a right join between the two datasets.

head(ice\_cream)

## key name  
## 1 0\_talenti ALPHONSO MANGO SORBETTO  
## 2 1\_talenti BANANA CARAMEL CRUNCH  
## 3 2\_talenti BELGIAN CHOCOLATE GELATO  
## 4 3\_talenti BLACK RASPBERRY CHOCOLATE CHIP GELATO  
## 5 4\_talenti BLACK RASPBERRY VANILLA PARFAIT  
## 6 5\_talenti CARAMEL APPLE PIE GELATO  
## description  
## 1 Our simple and delicious Alphonso Mango Sorbetto is made with the finest ripe mangoes of Maharashtra, pure sugar, and fresh lemon. Itâ\200\231s the perfect dairy-free and vegan treat that everyone can enjoy.  
## 2 This flavor was inspired by a classic southern dessert: banana pudding. We started with two smooth and creamy layers of banana pudding gelato and included a layer of our signature Dulce de Leche. To top it off, we mixed in baked pie crust pieces for a unique Talenti twist.  
## 3 Our Belgian Chocolate gelato is made with melted, imported Belgian chocolate blended until smooth with fresh milk, cream, eggs and a hint of vanilla. When it comes to chocolate gelato, thereâ\200\231s good, thereâ\200\231s better, and thereâ\200\231s Belgian.  
## 4 Tart and sweet, our Black Raspberry Chocolate Chip gelato combines black raspberries (from a lovely Oregon father/daughter-run raspberry farm, to boot) with rich chocolatey chips, all folded into a creamy gelato.  
## 5 This Layer was inspired by one of our favorite desserts â\200“ parfaits. Itâ\200\231s well-balanced and has a warm and satisfying taste that we are obsessed with. We start with Black Raspberry gelato which is crafted with black raspberry puree that comes from a farm in the Pacific Northwest. Next, we add oat crisps (which are similar to a crunchy oatmeal cookie) and a third layer of Blueberry sauce which is made from all-American blueberries. Finally we add a fourth layer of vanilla gelato and a fifth layer of oat crisps. We think this Black Raspberry Vanilla Parfait is the perfect dessert!  
## 6 We start by steeping real cinnamon sticks in fresh milk and cream to make our delectable cinnamon gelato, then blend in bits of flaky piecrust, apple pieces, and a caramel swirl to create this one-of-a-kind Caramel Apple Pie gelato. Our methods might be Italian, but this recipe clearly leans American. This is special batch gelato, meaning itâ\200\231s only available for a few months when itâ\200\231s in season!  
## rating rating\_count  
## 1 4.7 139  
## 2 4.2 25  
## 3 4.8 27  
## 4 4.0 105  
## 5 4.5 65  
## 6 4.8 115  
## ingredients  
## 1 MANGOS, WATER, SUGAR, DEXTROSE, LEMON JUICE, CAROB BEAN GUM  
## 2 MILK, BANANAS, SUGARâ\200 , WHEAT FLOUR, CREAM, NONFAT MILK, EGG YOLKS, WATER, EGG WHITES, COCONUT OIL, CORN SYRUPâ\200 , LEMON JUICE, WHOLE EGGS, CAROB BEAN GUM, BAKING SODA, VANILLA EXTRACT, NATURAL FLAVOR, SALT  
## 3 MILK, SUGAR, CREAM, EGG AND EGG YOLK, DEXTROSE, CHOCOLATE, COCOA (PROCESSED WITH ALKALI), COCOA BUTTER, CHOCOLATE, CAROB GUM, SOY LECITHIN, NATURAL FLAVOR  
## 4 MILK, SUGAR, CREAM, BLACK RASPBERRIES, DEXTROSE, CHOCOLATE (PROCESSED WITH ALKALI), COCONUT OIL, CAROB GUM, SOYBEAN OIL, VANILLA EXTRACT, SOY LECITHIN NATURAL FLAVOR  
## 5 MILK, SUGAR, CREAM, BLACK RASPBERRY PUREE, ROLLED OATS, COCONUT OIL, WHEAT FLOUR, WATER, TAPIOCA SYRUP, BROWN SUGAR, BLUEBERRIES, DEXTROSE, NATURAL FLAVOR, SUNFLOWER LECITHIN, CAROB BEAN GUM, GUAR GUM, VANILLA EXTRACT, PECTIN, RED CABBAGE JUICE CONCENTRATE (FOR COLOR), SALT  
## 6 MILK, SUGAR, CREAM, DEXTROSE, SKIM MILK, WHEAT FLOUR, CINNAMON, DRIED APPLES, CORN SYRUP, VANILLA EXTRACT, CAROB GUM, COCONUT OIL, BAKING SODA, EGGS, SALT

head(review)

## key author date stars  
## 1 0\_talenti amys 2018-06-22 5  
## 2 0\_talenti Hesskr 2017-10-29 5  
## 3 0\_talenti kookie 2018-09-20 5  
## 4 0\_talenti mike99999 2017-09-25 1  
## 5 0\_talenti JaiCandy 2016-06-23 5  
## 6 0\_talenti SR Horsemanship 2017-09-25 5  
## title helpful\_yes helpful\_no  
## 1 Mango Madness 12 0  
## 2 Mangos Mashed Up 21 0  
## 3 Surprised by the flavor 6 0  
## 4 dont need limon, just distorting the mango flavor 21 45  
## 5 AMAZING 18 0  
## 6 Absolutely delicious! 26 1  
## text  
## 1 Yâ\200\231all! Iâ\200\231m sick as a dog!!! For reals! Temp of over 101. Begged my mom to get me some talenti because itâ\200\231s my new favorite again. Anyways I told her to get something non-dairy for my throat so it wouldnâ\200\231t get all yucky. She brings me this flavor.\nO.M.G. THIS IS SO GOOD!!!!!\nOk, I have a cold so everything tastes bland so this shouldnâ\200\231t be all that great right? No. This tastes like the actual mango, but like better than the actual mango!!!\nSeriously! Every flavor I try is my new favorite!! I 100% recommend this and all the other flavors! Talenti is the best. Oh! And thereâ\200\231s always a treat at the end. You can reuse the container. These guys are the best!!  
## 2 This is seriously like eating fresh mangoes smashed up and put into a container. Many other surveys on the market remind me to much of ice cream and have a creaminess to them. Some people enjoy this, however I prefer a sorbet to taste just like fresh fruit. The ingredients in this could not be more direct. Mangoes, sugar, water, dextrose And lemon juice. At first I did not think I would like the tartness from the lemon juice. However it gives just a bit of lemon flavor, Which almost accentuates the mangoes. This is by far our favorite!  
## 3 When I first tried this sorbetto (the first thing I ever tried from Talenti), I wasn't expecting much. I was thinking it would probably taste like the other mango sorbets that I had tried, enough flavor to know that it was mango but not much else. Then, I had my first spoonful. To me, it tasted as though someone had pureed and frozen a mango and put it in the container. It was delicious. The flavor was so intense. I've convinced several other family members to try it since then, and they've agreed with me. We've all become Talenti converts thanks to this mango sorbetto.  
## 4 ice cream, do not need limon, just distorting the mango flavor  
## 5 Really taste like you're eating a chilled mango. It is one of my new favorite flavors. It is on the sweeter side though. Would love to see it with less sugar, so you can really taste the mango more. Otherwise its amazing!  
## 6 Perfect just the way it is.\nI LOVE IT! New favorite.

unique(ice\_cream$name)

## [1] "ALPHONSO MANGO SORBETTO"   
## [2] "BANANA CARAMEL CRUNCH"   
## [3] "BELGIAN CHOCOLATE GELATO"   
## [4] "BLACK RASPBERRY CHOCOLATE CHIP GELATO"  
## [5] "BLACK RASPBERRY VANILLA PARFAIT"   
## [6] "CARAMEL APPLE PIE GELATO"   
## [7] "CARAMEL COOKIE CRUNCH GELATO"   
## [8] "CARIBBEAN COCONUT GELATO"   
## [9] "CHOCOLATE CHERRY CHEESECAKE"   
## [10] "CHOCOLATE CHIP COOKIE DOUGH GELATO"   
## [11] "CHOCOLATE PEANUT BUTTER CUP GELATO"   
## [12] "COCONUT ALMOND CHOCOLATE GELATO"   
## [13] "COCONUT CHOCOLATE COOKIE"   
## [14] "COFFEE CHOCOLATE CHIP GELATO"   
## [15] "COFFEE COOKIE CRUMBLE"   
## [16] "COLD BREW COFFEE SORBETTO"   
## [17] "DARK CHOCOLATE CHERRY"   
## [18] "DARK CHOCOLATE SORBETTO"   
## [19] "DOUBLE DARK CHOCOLATE GELATO"   
## [20] "FUDGE BROWNIE GELATO"   
## [21] "HAZELNUT CHOCOLATE CHIP GELATO"   
## [22] "KEY LIME PIE GELATO"   
## [23] "LEMON BERRY PIE"   
## [24] "MADAGASCAN VANILLA BEAN GELATO"   
## [25] "MEDITERRANEAN MINT GELATO"   
## [26] "MINT FUDGE COOKIE"   
## [27] "OLD WORLD EGGNOG GELATO"   
## [28] "ORGANIC BROWN BUTTER CARAMEL"   
## [29] "ORGANIC CHOCOLATE MOUSSE GELATO"   
## [30] "ORGANIC GINGER MATCHA GELATO"   
## [31] "ORGANIC OAK-AGED VANILLA GELATO"   
## [32] "PACIFIC COAST PISTACHIO GELATO"   
## [33] "PEANUT BUTTER FUDGE SORBETTO"   
## [34] "PEANUT BUTTER VANILLA FUDGE"   
## [35] "PEPPERMINT BARK GELATO"   
## [36] "PUMPKIN PIE"   
## [37] "RASPBERRY CHEESECAKE GELATO"   
## [38] "ROMAN RASPBERRY SORBETTO"   
## [39] "SALTED CARAMEL TRUFFLE"   
## [40] "SEA SALT CARAMEL GELATO"   
## [41] "SOUTHERN BUTTER PECAN GELATO"   
## [42] "STRAWBERRY HIBISCUS SORBETTO"   
## [43] "VANILLA BLUEBERRY CRUMBLE GELATO"   
## [44] "VANILLA CARAMEL SWIRL GELATO"   
## [45] "VANILLA FUDGE COOKIE"

nrow(review)

## [1] 4069

flavors <- ice\_cream[, c("key", "name")]  
review\_full <- right\_join(flavors, review, by="key")  
  
summary(ice\_cream)

## key name description rating   
## Length:45 Length:45 Length:45 Min. :2.800   
## Class :character Class :character Class :character 1st Qu.:4.200   
## Mode :character Mode :character Mode :character Median :4.400   
## Mean :4.356   
## 3rd Qu.:4.700   
## Max. :4.800   
## rating\_count ingredients   
## Min. : 13.00 Length:45   
## 1st Qu.: 34.00 Class :character   
## Median : 72.00 Mode :character   
## Mean : 90.53   
## 3rd Qu.:115.00   
## Max. :334.00

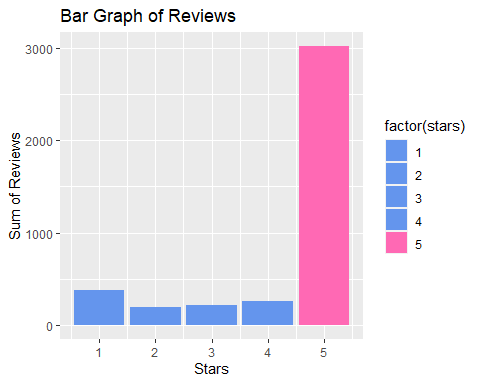
summary(review)

## key author date stars   
## Length:4069 Length:4069 Length:4069 Min. :1.000   
## Class :character Class :character Class :character 1st Qu.:4.000   
## Mode :character Mode :character Mode :character Median :5.000   
## Mean :4.312   
## 3rd Qu.:5.000   
## Max. :5.000   
## title helpful\_yes helpful\_no text   
## Length:4069 Min. : 0.000 Min. : 0.0000 Length:4069   
## Class :character 1st Qu.: 0.000 1st Qu.: 0.0000 Class :character   
## Mode :character Median : 0.000 Median : 0.0000 Mode :character   
## Mean : 1.283 Mean : 0.5869   
## 3rd Qu.: 1.000 3rd Qu.: 0.0000   
## Max. :84.000 Max. :98.0000

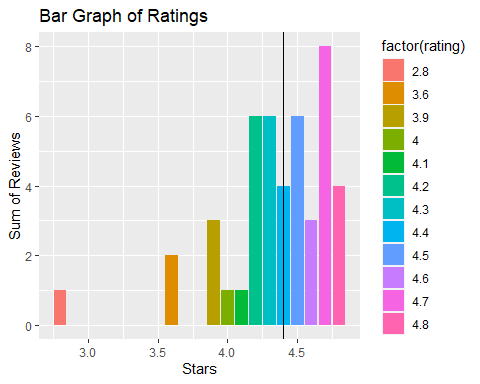
### **Visualizations**

The graphs below show that there are mostly 5-star reviews in the review dataset, with a median rating of 4.4. Reviews are not evenly distributed by flavor, meaning some flavors have a lot more reviews than others. Finally, since my goal is to evaluate what people like most about top flavors using reviews, I created a boxplot (and barplot) for unhelpful reviews. While there are outliers, most reviews have minimal downvotes. Later on, I’ll remove some high-downvote outliers for more accurate sentiment analysis.

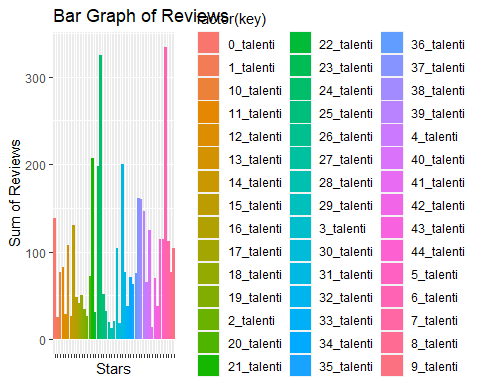
library(ggplot2)  
  
ggplot(data = review, aes(x = stars, fill = factor(stars))) +  
 geom\_bar() +  
 scale\_fill\_manual(values = c("cornflowerblue", "cornflowerblue", "cornflowerblue", "cornflowerblue", "hotpink")) +  
 labs(x = "Stars", y = "Sum of Reviews", title = "Bar Graph of Reviews")



ggplot(data = ice\_cream) +  
 geom\_bar(mapping = aes(x = rating, fill = factor(rating))) +  
 geom\_vline(xintercept = median(ice\_cream$rating)) +  
 labs(x = "Stars", y = "Sum of Reviews", title = "Bar Graph of Ratings")



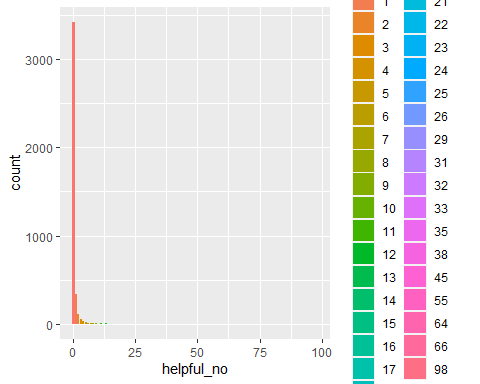
ggplot(data = review\_full) +  
 geom\_bar(mapping = aes(x = key, fill = factor(key))) +  
 labs(x = "Stars", y = "Sum of Reviews", title = "Bar Graph of Reviews") +  
 theme(axis.text.x = element\_blank())



ggplot(data = review, aes(x = helpful\_no)) +   
 geom\_boxplot()

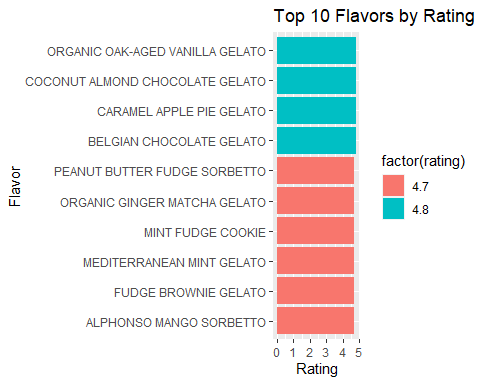


ggplot(data = review, aes(x = helpful\_no)) +   
 geom\_bar(mapping = aes(x = helpful\_no, fill = factor(helpful\_no)))

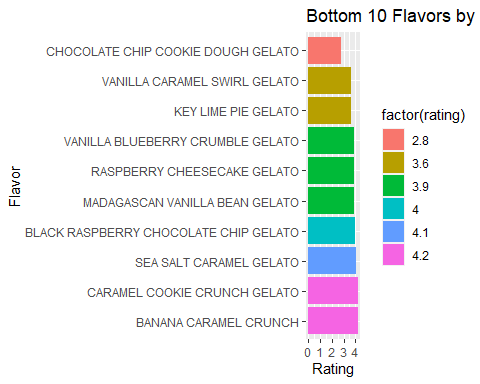


Below are the Top 10 and Bottom 10 Flavors by Rating, given in the review dataset. Bottom 10 is sorted in reverse so that the worst ranked flavor, Chocolate Chip Cookie Dough Gelato, is shown at the top.

top\_10 <- head(ice\_cream[order(ice\_cream$rating, decreasing=TRUE), ], 10)  
ggplot(data = top\_10) +  
 geom\_bar(mapping = aes(x = rating, y = reorder(name, rating),   
 fill = factor(rating)), stat = "identity") +  
 labs(x = "Rating", y = "Flavor", title = "Top 10 Flavors by Rating")



bottom\_10 <- head(ice\_cream[order(ice\_cream$rating, decreasing=FALSE), ], 10)  
ggplot(data = bottom\_10) +  
 geom\_bar(mapping = aes(x = rating, y = reorder(name, -rating),   
 fill = factor(rating)), stat = "identity") +  
 labs(x = "Rating", y = "Flavor", title = "Bottom 10 Flavors by Rating")



# **Sentiment Analysis**

### **Function Definitions**

I created some useful functions below. One cleans the text by removing all non-letter characters. Two convert reviews into relevant word components. One calculates the percentage of positive (and negative) sentiment words across all reviews by flavor.

#cleans text (removes all non-letter characters)  
clean\_text <- function(review\_df) {  
 text\_list <- character(0)  
 for (i in review\_df$text) {  
 #get rid of anything that isn't uppercase/lowercase letter  
 i <- str\_replace\_all(i, "[^a-zA-Z]", " ")  
 text\_list <- append(text\_list, i)  
 }  
 review\_df$text <- text\_list  
 return(review\_df)  
   
}  
  
  
#converts reviews for specific flavor into words  
review\_to\_word <- function(review\_df, name) {  
 review\_flavor <- review\_df[review\_df$name == name,]  
 clean\_reviews <- clean\_text(review\_flavor)  
 review\_in\_words <- clean\_reviews %>% mutate(text = tolower(text)) %>% unnest\_tokens(input=text, output=word) %>% anti\_join(stop\_words, by = "word") %>% count(word)  
 return(review\_in\_words)  
}  
  
#converts ALL reviews into words  
review\_to\_word\_all <- function(review\_df) {  
 clean\_reviews <- clean\_text(review\_full)  
 review\_in\_words <- clean\_reviews %>% mutate(text = tolower(text)) %>% unnest\_tokens(input=text, output=word) %>% anti\_join(stop\_words, by = "word") %>% group\_by(name)  
 return(review\_in\_words)  
}  
  
get\_sentiment\_percents <- function(review\_words, sentiments\_in) {  
 sentiment\_counts <- review\_words %>% inner\_join(sentiments\_in, by = "word") %>% group\_by(name) %>% count(sentiment)  
 sentiment\_sums <- sentiment\_counts %>% group\_by(name) %>% summarise(total\_count = sum(n))  
 sentiment\_percents <- sentiment\_counts %>% full\_join(sentiment\_sums, by = "name") %>% mutate(percent\_each = n / total\_count)  
 return(sentiment\_percents)  
}

### **Sentiment Processing**

Since I knew that high downvote reviews were outliers, I knew there probably weren’t many of them. Thus, for efficiency, I manually looked at the ratio between upvotes and downvotes. I did this to make sure I wasn’t eliminating reviews that simply had a high amount of vote interaction (high downvotes and even higher upvotes). All of the reviews had more downvotes than upvotes, so I eliminated them from analysis. I commented these lines of code out so large dataframes wouldn’t be printed in the final review.

I used “bing” sentiments, which is a dictionary-style library of words with binary definitions indicating “positive” or “negative” sentiment. I inner-joined the “bing” dictionary with the words in the reviews, which labeled the cleaned-up review words as positive or negative. I did this in one of the above functions (get\_sentiment\_percents). The below code chunk calls these functions and stores the percentage data in sentiment\_percents.

There are some limitations to using bing definitions. AI has limitations in interpreting context and/or human communication when classifying as strictly “positive” or “negative.” As an example, in the context of ice cream, a person saying they are “addicted” is probably meant facetiously as a good thing. However, the computer will interpret this as a negative word. Important to note is that traits such as “lemon,” “crumble,” and “dark” are also treated as negatives, which is a drawback when using this type of analysis.

Keeping limitations and the project goal in mind, the technique remains a time-efficient and reasonably effective method. It digests a large amount of review text into words and facilitates easy visualizations for gaining data insight. Looking at ratings data alongside sentiment analysis data will help ensure insights are accurate/the technique is effective. Likewise, looking at sentiment analysis data alongside ratings data will help ensure the most insightful positive descriptors are highlighted, something the ratings data couldn’t do as well by itself.

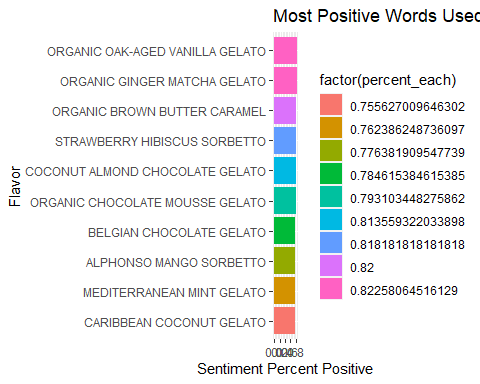
#review\_full[review\_full$helpful\_no >= 25,]  
#review\_full <- review\_full[review\_full$helpful\_no < 25,]  
  
sentiments <- get\_sentiments("bing")  
  
ratings\_text <- tibble(text = review\_full$text, name = review\_full$name)  
ratings\_text <- clean\_text(ratings\_text)  
ratings\_text\_full <- review\_to\_word\_all(ratings\_text)  
sentiment\_percents <- get\_sentiment\_percents(ratings\_text\_full, sentiments)

### **Highest and Lowest Percentage Positive Sentiment**

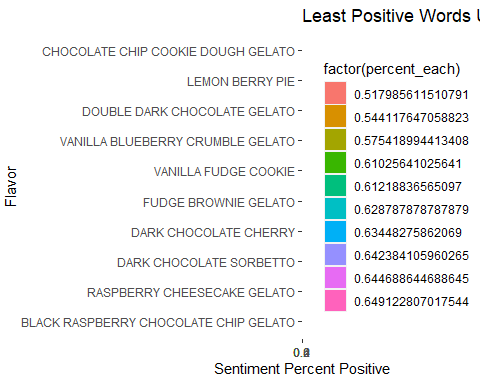
The below bar graphs show which flavors received the highest and lowest percentages of positive sentiment words in their reviews. These graphs help provide supplementary insight to ratings data, as the highest-rated flavor might have vaguer/less insightful reviews than another top-rated flavor.

A reasonable amount of similarity with the original Top 10 is shown. Organic Oak-Aged Vanilla Gelato remains the top flavor, Chocolate Chip Cookie Dough Gelato remains the worst, and no flavor from the Bottom 10 Ratings shows up in the Top 10 Sentiment-Ratings (and vice-versa). Many flavors from the Top 10 Ratings show up in different places of the Top 10 Sentiment-Ratings, including Alphonso Mango Sorbetto and Mediterranean Mint Gelato. Likewise, some from the Bottom 10 Ratings show up in the Bottom 10 Sentiment-Ratings, including Vanilla Blueberry Crumble Gelato and Raspberry Cheesecake Gelato.

sentiment\_percents <- ungroup(sentiment\_percents)  
sentiment\_positive <- sentiment\_percents[sentiment\_percents$sentiment == "positive",]  
top\_10\_sentiments <- head(sentiment\_positive[order(sentiment\_positive$percent\_each, decreasing=TRUE), ], 10)  
ggplot(data = top\_10\_sentiments) +   
 geom\_bar(mapping = aes(x = percent\_each, y = reorder(name, percent\_each), fill = factor(percent\_each)), stat="identity") +   
 labs(x = "Sentiment Percent Positive", y = "Flavor", title = "Most Positive Words Used in Reviews")



sentiment\_negative <- sentiment\_percents[sentiment\_percents$sentiment == "positive",]  
top\_10\_sentiments <- tail(sentiment\_negative[order(sentiment\_positive$percent\_each, decreasing=TRUE), ], 10)  
ggplot(data = top\_10\_sentiments) +   
 geom\_bar(mapping = aes(x = percent\_each, y = reorder(name, -percent\_each), fill = factor(percent\_each)), stat="identity") +   
 labs(x = "Sentiment Percent Positive", y = "Flavor", title = "Least Positive Words Used in Reviews")



# **Features in Sentiment Analysis, by (select) Flavors**

### **Function Definitions**

I created some more functions for flavor-specific attributes. One retrieves positive words from reviews, and another retrieves negative words. Two find reviews with 3+ star ratings and 1-2 star ratings, respectively. The final code chunk contains four functions to print word clouds. I needed a workaround solution to have all word clouds show up in the Knit HTML R-Markdown file, so I got help from this URL: <https://stackoverflow.com/questions/59443054/wordcloud-error-in-rmarkdown-only-one-wordcloud-shows-up-in-html>

get\_positive\_words <- function(flavor, sentiments\_in) {  
 flavor\_reviews\_df <- review\_full[review\_full$name == flavor,]  
 flavor\_reviews\_df <- flavor\_reviews\_df[flavor\_reviews\_df$stars >= 3,]  
 flavor\_reviews\_df <- clean\_text(flavor\_reviews\_df)  
 flavor\_reviews\_df <- review\_to\_word(flavor\_reviews\_df, flavor)  
   
 flavor\_reviews\_df <- flavor\_reviews\_df %>% inner\_join(sentiments\_in, by = "word")  
 flavor\_reviews\_df <- flavor\_reviews\_df[flavor\_reviews\_df$sentiment == "positive",]  
   
 return(flavor\_reviews\_df)  
}  
  
get\_negative\_words <- function(flavor, sentiments\_in) {  
 flavor\_reviews\_df <- review\_full[review\_full$name == flavor,]  
 flavor\_reviews\_df <- clean\_text(flavor\_reviews\_df)  
 flavor\_reviews\_df <- review\_to\_word(flavor\_reviews\_df, flavor)  
   
 flavor\_reviews\_df <- flavor\_reviews\_df %>% inner\_join(sentiments\_in, by = "word")  
 flavor\_reviews\_df <- flavor\_reviews\_df[flavor\_reviews\_df$sentiment == "negative",]  
   
 return(flavor\_reviews\_df)  
}

isolate\_flavor <- function(flavor) {  
 isolated\_flavor\_df <- review\_full[review\_full$name == flavor,]  
 return(isolated\_flavor\_df)  
}  
  
three\_and\_above <- function(flavor) {  
 flavor\_df <- isolate\_flavor(flavor)  
 three\_and\_above\_df <- flavor\_df[flavor\_df$stars >= 3,]  
 return(three\_and\_above\_df)  
}  
  
three\_and\_below <- function(flavor) {  
 flavor\_df <- isolate\_flavor(flavor)  
 three\_and\_below\_df <- flavor\_df[flavor\_df$stars < 3,]  
 return(three\_and\_below\_df)  
}

suppressWarnings({  
 library(webshot)  
 library(wordcloud2)  
 library(htmlwidgets)  
})  
#https://stackoverflow.com/questions/59443054/wordcloud-error-in-rmarkdown-only-one-wordcloud-shows-up-in-html  
#I looked at above page for screenshot workaround  
  
#webshot::install\_phantomjs()  
  
print\_three\_and\_above <- function(flavor, size\_in) {  
 three\_and\_above\_df <- three\_and\_above(flavor)  
 three\_and\_above\_df <- clean\_text(three\_and\_above\_df)  
 three\_and\_above\_df <- review\_to\_word(three\_and\_above\_df, flavor)  
   
 #MOST COMMON WORDS IN ALL REVIEWS, THREE-STAR PLUS  
 print(wordcloud2(three\_and\_above\_df, size = size\_in, color ="random-light", backgroundColor = "dark"))  
   
 #screenshot workaround (see above)  
 html\_word\_cloud <- wordcloud2(three\_and\_above\_df, size = size\_in, color = "random-light", backgroundColor = "dark")  
saveWidget(html\_word\_cloud, '1.html', selfcontained = F)  
webshot('1.html', '1.png', vwidth=700,vheight=500, delay = 5)  
}  
  
print\_positive\_words <- function(flavor, size\_in) {  
 positive\_words\_df <- get\_positive\_words(flavor, sentiments)  
 #MOST COMMON POSITIVE WORDS IN ALL REVIEWS  
 print(wordcloud2(positive\_words\_df, size = size\_in, color = "random-light", backgroundColor = "dark"))  
   
#screenshot workaround (see above)  
html\_word\_cloud <- wordcloud2(positive\_words\_df, size = size\_in, color = "random-light", backgroundColor = "dark")  
saveWidget(html\_word\_cloud, '1.html', selfcontained = F)  
webshot('1.html', '1.png', vwidth=700,vheight=500, delay = 5)  
  
}  
  
print\_three\_and\_below <- function(flavor, size\_in) {  
 three\_and\_below\_df <- three\_and\_below(flavor)  
 three\_and\_below\_df <- clean\_text(three\_and\_below\_df)  
 three\_and\_below\_df <- review\_to\_word(three\_and\_below\_df, flavor)  
   
 #MOST COMMON WORDS IN ALL REVIEWS, TWO AND ONE STAR  
 print(wordcloud2(three\_and\_below\_df, size = size\_in, color ="random-light", backgroundColor = "black"))  
   
#screenshot workaround (see above)  
html\_word\_cloud <- wordcloud2(three\_and\_below\_df, size = size\_in, color = "random-light", backgroundColor = "dark")  
saveWidget(html\_word\_cloud, '1.html', selfcontained = F)  
webshot('1.html', '1.png', vwidth=700,vheight=500, delay = 5)  
}  
  
  
print\_negative\_words <- function(flavor, size\_in) {  
 negative\_words\_df <- get\_negative\_words(flavor, sentiments)  
 #MOST COMMON NEGATIVE WORDS IN ALL REVIEWS  
 print(wordcloud2(negative\_words\_df, size = size\_in, color = "random-light", backgroundColor = "black"))  
   
 #screenshot workaround (see above)  
 html\_word\_cloud <- wordcloud2(negative\_words\_df, size = size\_in, color = "random-light", backgroundColor = "dark")  
saveWidget(html\_word\_cloud, '1.html', selfcontained = F)  
webshot('1.html', '1.png', vwidth=700,vheight=500, delay = 5)  
  
}

### **Positive Product Features by Flavor**

I selected a handful of Talenti’s gelatos and sorbettos that showed up in both the Highly Rated and Highly Sentiment-Rated Top 10 lists. Below shows an up-close look at reviews from these select flavors.

The first word cloud shows the most commonly used words in reviews with 3+ star ratings. This is useful to look at because it will show non-sentiment descriptors (e.g.: chocolate, pint) that are still relevant. In my write-up, I’ll call this the “three-and-above word cloud.” The next word cloud shows positive words from these reviews. This is useful because it highlights the sentimental descriptors for us, which is much better than us digging around for them in the three-and-above word cloud. I’ll call this the “positive word cloud.”

## **ORGANIC OAK-AGED VANILLA GELATO**

### **Tied for First Highest Rated, Tied for Highest Sentiment Analysis**

### **Flavor Description**

vanilla <- ice\_cream[ice\_cream$name == "ORGANIC OAK-AGED VANILLA GELATO",]  
vanilla$description <- str\_replace\_all(vanilla$description, "â€™", "'")  
vanilla$description <- str\_replace\_all(vanilla$description, "â€”", ", ")  
print(vanilla$description)

## [1] "Some things get better with age. Like our new Organic Oak-Aged Vanilla gelato, the mature, more sophisticated cousin of our classic Madagascan Vanilla. It's an elevated gelato crafted with Organic bourbon Vanilla and hints of oak. It's a taste you'll want to take your time with, but might not be able to."

print(vanilla$ingredients)

## [1] "INGREDIENTS: ORGANIC SKIM MILK, ORGANIC CREAM, ORGANIC CANE SUGAR, ORGANIC DEXTROSE, OAKWOOD EXTRACT, ORGANIC CAROB BEAN GUM, ORGANIC VANILLA EXTRACT"

### **ANALYSIS**

The three-and-above word cloud shows the word “aged” being used 5 times, a relatively high amount, along with “cream.” Words such as “notes,” “brandy,” “bourbon,” “alcohol,” and “spirits” appear in this cloud as well, likely indicating positive sentiment towards the “bourbon Vanilla” noted in the flavor description. “Complex” also makes an appearance, along with “elegant.” Many other desserts/food items show up in the word cloud, including “coconut,” “chocolate,” “custard,” “berries,” and “pecans.”

“Perfectly,” “balanced,” “fine,” and “smooth” are words that are relatively often used in reviews for the Organic Oak-Aged Vanilla, as shown in the positive word cloud. “Sophisticated” and “fancy” show up, along with “complement.”

Based on the word clouds, people who like this gelato are probably happiest with its “sophisticated” flavor, including the notes of bourbon, its smooth and creamy texture, and its balanced quality. People may also enjoy pairing this flavor with other desserts or food items. Analysis of this flavor may be limited due to a low number of reviews.

print\_three\_and\_above("ORGANIC OAK-AGED VANILLA GELATO", 1.5)



print\_positive\_words("ORGANIC OAK-AGED VANILLA GELATO", 1)



## **COCONUT ALMOND CHOCOLATE GELATO**

### **Tied for First Highest Rated, Top 10 Sentiment Analysis**

### **Flavor Description**

cacg <- ice\_cream[ice\_cream$name == "COCONUT ALMOND CHOCOLATE GELATO",]  
print(cacg$description)

## [1] "This flavor is like if Caribbean Coconut Talenti took its best friends on vacation. Coconut shreds, semi-sweet chocolatey flakes, and salted roasted almonds make quite a party."

print(cacg$ingredients)

## [1] "MILK, SUGAR, CREAM, CHOCOLATE (PROCESSED WITH ALKALI), ALMONDS, COCONUT CONCENTRATE, COCONUT, DEXTROSE, COCONUT AND SOYBEAN AND SUNFLOWER OILS, SALT, CAROB BEAN GUM, SOY LECITHIN, VANILLA EXTRACT, NATURAL FLAVOR"

### **ANALYSIS**

The three-and-above word cloud shows very large sizes for “coconut,” “almond(s),” chocolate,” “ice,” and “cream.” Reasonably large in this cloud are “combination,” “hard,” “blend,” “taste,” “creamy,” and “texture.”

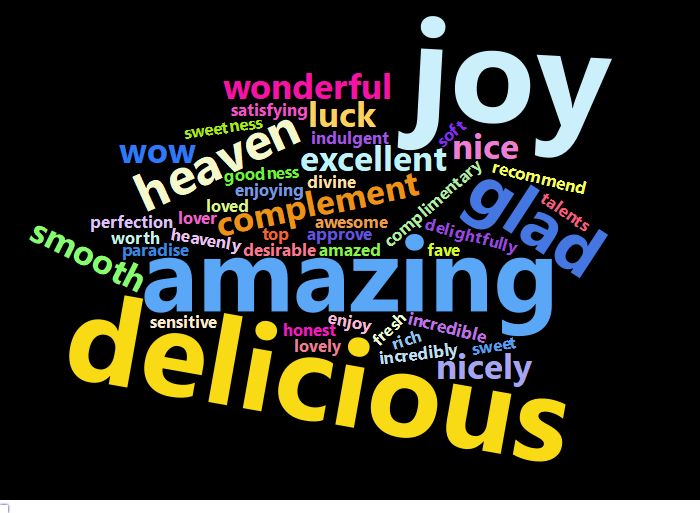
Two word clouds are printed for positive words, one to zoom in (the second printed) and one as a full-size picture (the third printed). This is because two words, “love” and “favorite,” significantly outsize the other words in the cloud. “Smooth” and “nicely” make appearances in this word cloud. Less-used yet similar words in the zoomed-in cloud include “sweet(ness),” “indulgent,” “rich,” and “complement.”

The Coconut Almond Chocolate Gelato’s flavor combination is likely what people like most. Some of the largest words in the clouds were ingredients, synonyms for combinations, and descriptors for the taste of the gelato. Similar to the Organic Oak-Aged Vanilla, reviews also commonly used “smooth,” “creamy,” and “texture,” indicating positive sentiment towards the gelato’s texture. “Favorite” was a very large word in the positive word cloud, which shows that this flavor is important or special to some customers. It could be attracting new customers and/or reliable sales, if this product provides something so important/unique to the frozen dessert market that customers keep buying it.

print\_three\_and\_above("COCONUT ALMOND CHOCOLATE GELATO", 1.5)



print\_positive\_words("COCONUT ALMOND CHOCOLATE GELATO", 3)



print\_positive\_words("COCONUT ALMOND CHOCOLATE GELATO", 1)



## **CARAMEL APPLE PIE GELATO**

### **Tied for First Highest Rated**

capg <- ice\_cream[ice\_cream$name == "CARAMEL APPLE PIE GELATO",]  
capg$description <- str\_replace\_all(capg$description, "â€™", "'")  
print(capg$description)

## [1] "We start by steeping real cinnamon sticks in fresh milk and cream to make our delectable cinnamon gelato, then blend in bits of flaky piecrust, apple pieces, and a caramel swirl to create this one-of-a-kind Caramel Apple Pie gelato. Our methods might be Italian, but this recipe clearly leans American. This is special batch gelato, meaning it's only available for a few months when it's in season!"

print(capg$ingredients)

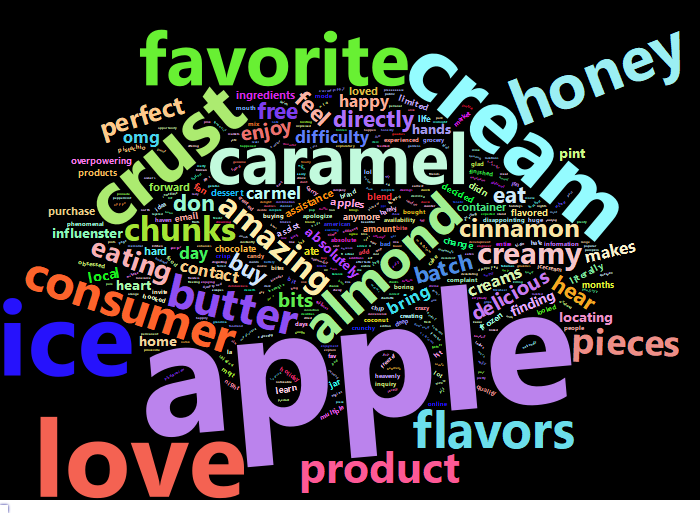
## [1] "MILK, SUGAR, CREAM, DEXTROSE, SKIM MILK, WHEAT FLOUR, CINNAMON, DRIED APPLES, CORN SYRUP, VANILLA EXTRACT, CAROB GUM, COCONUT OIL, BAKING SODA, EGGS, SALT"

### **ANALYSIS**

“Apple,” “caramel,” “honey,” “cream,” “crust,” “pieces,” and “flavors” are some of the largest words in the three-and-above word cloud. “Seasonal,” “product,” and “services” also make appearances. “Sweet,” “sweetness,” and “smooth” show up in the positive word cloud, along with “favorite,” which is one of the largest.

Similar to the Coconut Almond Chocolate Gelato, people who like this gelato probably like the flavor and possibly the seasonal component. They also probably like it for its sweetness and texture. It is also a “favorite” among some people.

print\_three\_and\_above("CARAMEL APPLE PIE GELATO", 1.2)



print\_positive\_words("CARAMEL APPLE PIE GELATO", 2.5)



print\_positive\_words("CARAMEL APPLE PIE GELATO", 1)



## **ALPHONSO MANGO SORBETTO**

### **Tied for 2nd Highest Rated, Top Sentiment-Rated**

ams <- ice\_cream[ice\_cream$name == "ALPHONSO MANGO SORBETTO",]  
ams$description <- str\_replace\_all(ams$description, "â€™", "'")  
print(ams$description)

## [1] "Our simple and delicious Alphonso Mango Sorbetto is made with the finest ripe mangoes of Maharashtra, pure sugar, and fresh lemon. It's the perfect dairy-free and vegan treat that everyone can enjoy."

print(ams$ingredients)

## [1] "MANGOS, WATER, SUGAR, DEXTROSE, LEMON JUICE, CAROB BEAN GUM"

### **ANALYSIS**

The three-and-above word cloud shows frequent usage of the words “favorite,” “smooth,” “sweet,” and “mangoes,” along with “dairy,” “vegan,” “sorbet,” and “ingredients.” The positive word cloud cements these sentiments and adds on the words “fresh” and “refreshing.”

This suggests the Alphonso Mango Sorbetto may be popular among a crowd looking for dairy-free options. People may also like that it is “fresh” and “refreshing,” which makes sense for a fruit-based sorbet.

print\_three\_and\_above("ALPHONSO MANGO SORBETTO", 1.5)



print\_positive\_words("ALPHONSO MANGO SORBETTO", 1.5)



## **PEANUT BUTTER FUDGE SORBETTO**

### **Tied for 2nd Highest Rated**

### **Flavor Description**

test <- ice\_cream[ice\_cream$name == "PEANUT BUTTER FUDGE SORBETTO",]  
test$description <- str\_replace\_all(test$description, "â€™", "'")  
test$description <- str\_replace\_all(test$description, "â€œ", "'")  
test$description <- str\_replace\_all(test$description, "â€\u009d", "'")  
print(test$description)

## [1] "Our rich, creamy, vegan Peanut Butter Fudge sorbetto features a decadent swirl of dairy-free fudge. Don't be surprised if you find yourself asking 'How did my peanut butter find its way to the freezer? And how does it taste this good?'"

print(test$ingredients)

## [1] "WATER, PEANUTS, SUGAR, DEXTROSE, SALT, COCOA PROCESSED WITH ALKALI, CORN SYRUP, COCONUT OIL, CHOCOLATE, CORN STARCH, SEA SALT, VANILLA EXTRACT, CAROB BEAN GUM, GUAR GUM, XANTHAN GUM"

### **ANALYSIS**

“Cream,” “creamy,” “fudge,” “chocolate,” and “flavor” are top words for this flavor, but the largest behind “butter” is “free.” “Milk,” “vegan,” “lactose,” and “allergy” make appearances in the three\_and\_above word cloud, with “vegan” being the largest out of the four. “Texture,” “thick,” and “rich” make appearances in this word cloud as well. Interestingly, “coconut” only shows up in the word cloud twice, and may refer to the coconut oil in the sorbet (which may be a dairy substitute). This could suggest the sorbet does not have a strong coconut flavor. More research would have to be done on the reviews mentioning coconut/why coconut oil is included in the sorbet to cement this idea.

“Free” is the largest word in the positive word cloud. This word cloud also shows “favorite,” “smooth,” and “sweet.”

Similar to above flavors, people who like this sorbet probably like the flavor and flavor combinations. However, more relevant to this flavor, people may like the vegan alternative it provides. Since many words related to the texture, such as “creamy” and “smooth,” people may like that the sorbet resembles a dairy-containing gelato or other frozen dessert.

print\_three\_and\_above("PEANUT BUTTER FUDGE SORBETTO", 1)



print\_positive\_words("PEANUT BUTTER FUDGE SORBETTO", 1)



## **CHOCOLATE CHIP COOKIE DOUGH GELATO and Other Low-Rated Flavors: A Quick Look**

### **ANALYSIS**

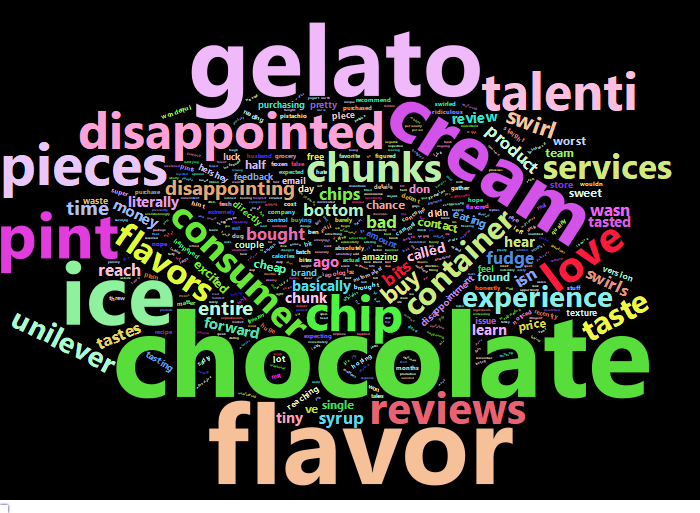
Finally, as a comparison, I took a look at the Chocolate Chip Cookie Dough Gelato, which had the worst ratings and worst sentiment-ratings.

“Chocolate,” “flavor,” “cream,” and “pieces” show up several times in the three\_and\_below word cloud, along with “chip,” “chunks,” “consumer,” and “disappointed.” “Container” and “pint” also make considerable appearances in the three\_and\_below word cloud. “Hard”/“frozen” show up, along with “disappointed,” “disappointment,” “concerns,” “false,” “lack,” “lacking,” “anomaly,” “inadequate,” “misleading,” “nonexistent,” “exaggeration,” and “overwhelmingly.” This suggests there might be a significant lack of a certain flavor or ingredient in containers/pints. Possibly, the “inadequate” ingredient might be the cookie dough pieces since “pieces” was in the word cloud and “hard” and “frozen” also show up in reviews.

For confirmation, I looked at the first few reviews and saw mostly complaints about a lack of/poor distribution of cookie dough pieces in pints.

Some reviews included frustration about this being a well-documented issue in reviews. Likewise, “services” and “reviews” also show up frequently in the three\_and\_below word cloud. It’s possible that increased customer engagement in this regard may help alleviate pressure on this product.

#cost, quality, balanced flavors (none overpowering)/alignment with product description, recipe changes,  
#customer engagement  
  
print\_three\_and\_below("CHOCOLATE CHIP COOKIE DOUGH GELATO", 1.7)



print\_negative\_words("CHOCOLATE CHIP COOKIE DOUGH GELATO", 1.5)

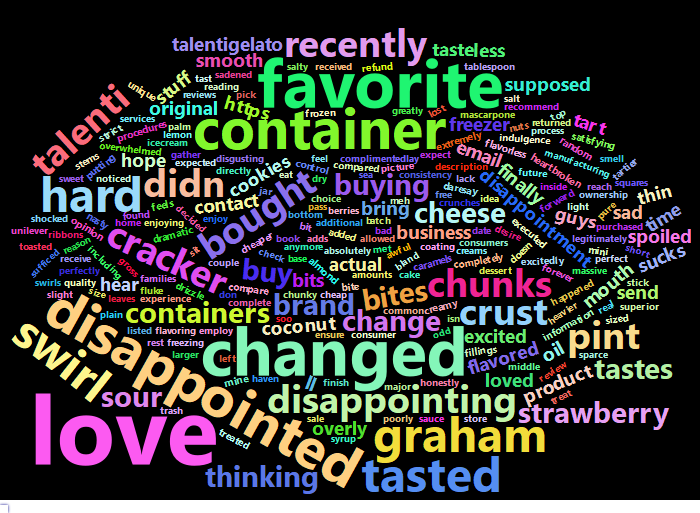


Below are word clouds for two more flavors, the Vanilla Blueberry Crumble Gelato and the Raspberry Cheesecake Gelato. **Only the negative words are printed for the former.** These both had a 3.9 average rating, placing them in the Bottom 10. Across all the word clouds, complaints seem to surround the quality of the product and imbalance of flavors. The Vanilla Blueberry Crumble had negative words such as “missed,” “mushy,” “overwhelming,” and “deceiving.” The Raspberry Cheesecake had negative words such as “hard,” “weird,” “chunky,” “lack,” “bland,” “sour,” and “spoiled.”

print\_negative\_words("VANILLA BLUEBERRY CRUMBLE GELATO", 2)



print\_three\_and\_below("RASPBERRY CHEESECAKE GELATO", 1.7)



print\_negative\_words("RASPBERRY CHEESECAKE GELATO", 1)

