Amazon Beauty Product Reviews: Sentiment Analysis & Web Application

STATS 418

Spring 2024

Anum Damani, Yuhuan He, Dani Wu, Cindy Xu

Table of Contents

Section I: Abstract

This project presents a comprehensive sentiment analysis using an Amazon Beauty product review dataset, created by Jianmo Ni, aiming to understand customer preferences and common product issues. Using this dataset of beauty product reviews, we applied Natural Language Processing techniques in Python to classify reviews into three categories: positive, negative, and neutral. We implemented an interactive dashboard using R Shiny that features several types of plots: polarity plots, time series plots, word clouds for verified versus unverified reviews, and word clouds for positive and negative sentiments. The overall polarity plot revealed that most Amazon reviews in the beauty category exhibit a slight positive sentiment, and it highlights patterns and trends in sentiment. The time series analysis shows an overall increase in positive sentiment over the years, with notable spikes around 2012 to 2013, and dips indicating periods of customer dissatisfaction. The positive-sentiment word clouds featured words related to bath and dental products, while the negative-sentiment word clouds featured words related to product quality. The user type word clouds, which create comparative word clouds from unverified or verified reviews, indicated that verified reviews showed more mixed sentiment, while unverified reviews showed a relatively positive sentiment. Our project faced several limitations; one of the biggest limitations was that our dataset documentation did not explain product IDs or names, which affected our ability to do product-specific analysis. We can make several improvements to the project in the future, including expanding our focus to include other product categories besides beauty, which would broaden the scope of our project. The link to our interactive dashboard and the link to our GitHub repository (containing all project files) are located in the Appendix.

Section II: About the Data

The Amazon Review dataset we selected for this project was from a GitHub website created by Jianmo Ni[1], a UC San Diego PhD student, who currently works at Google. We focused specifically on Amazon's "All Beauty" category. The "All Beauty" category includes products such as makeup, skin care, hair care, fragrance, etc.

Column Name	*Description*
overall Rating	given by the reviewer (1 to 5 stars)
verified	Indicates if the review is from a verified purchase
reviewerID	Unique identifier for the reviewer
productID	Unique identifier for the product
reviewerName	Name/Username of the reviewer
reviewText	The full text of the review
summaryReviewText	Short summary of the review
year	The year when the review was written
processed_reviewText	The preprocessed text of the review after tokenizing and lowercasing
filtered_reviewText	The filtered review after removing stopwords

Figure 1: Description of Columns in the Dataset

Section III: Data Cleaning & Preprocessing

We chose Python for data cleaning due to its powerful libraries and ease of use for various data preprocessing tasks such as tokenizing, lemmatizing, and removing stopwords. Below are detailed steps of our data cleaning and preprocessing:

1. The JSON data was transformed into a manageable, tabular format that was more suitable for analysis. This step included renaming columns to more meaningful names to ensure clarity of the variables.
2. We determined that there were no missing values in the data, and we decided to only keep columns that were necessary for our analysis.
3. In our original dataset, we had a column for the date on which the review was posted; the date column was in the following format: DD M, YYYY. We decided to extract only the year and stored it as a new variable called "year" to make our analysis more smooth.
4. Text data was preprocessed through tokenization, lowercasing, and removing stop words and punctuation. This step ensured the text data was in a consistent format for natural language processing tasks.
5. The stopwords list was extended to include beauty domain-specific words such as "product" and "used." These types of words may be irrelevant to the analysis, so by refining the data, we ensure that we only keep meaningful words.
6. Finally, the cleaned dataset was saved into a CSV file.

Section IV: Dashboard Methodology

While we used Python for sentiment analysis, we decided to use R Shiny to create our interactive dashboard. Although R can be a good choice for sentiment analysis, we chose Python due to the straightforward and popular “nltk” library, which was incredibly helpful. After completing the sentiment analysis, we felt most comfortable with implementing our plots in the R Shiny dashboard due to our familiarity with creating data visualizations in R.

The overall polarity plot helps identify overall trends in customer sentiment, highlighting extremes of positive and negative feedback. Users can choose how many reviews to view at a time. By visualizing sentiment polarity, users can better understand overall customer experiences for products and pinpoint areas that may need improvement. Additionally, the interactive nature of the plots allows users to explore the data in more depth, providing a dynamic tool for sentiment analysis and insights.

The time series plots in our dashboard provide a comprehensive view of sentiment trends over time, allowing users to observe how customer sentiment regarding beauty products has evolved. The plots include a general sentiment overview with aggregated positive and negative sentiments across all reviews, and users can select specific sentiments such as anger, anticipation, disgust, fear, joy, sadness, surprise, and trust for deeper insights into customer reactions. This feature is particularly helpful in identifying patterns and shifts in customer sentiment, highlighting periods of significant positive or negative feedback.

Word clouds are visual representations of textual data where the size of each word indicates the frequency it shows up or the overall importance of that word. The words that appear more often are displayed in much larger fonts whereas words that appear less often are shown in smaller fonts. This form of visualization is very useful in data analysis for many reasons, they give quick and easily digestible insights, identify key trends, and provide comparative analysis via a simple visual appeal.

Section V: Dashboard Plots

A. Overall Polarity Plot

The overall polarity plot was created over the entire dataset to visualize the sentiment distribution in the “All Beauty” data. Initially, the data was prepared by performing sentiment analysis on the review texts using R’s “syuzhet” library. Sentiment scores were calculated, and from these, polarity scores were derived by subtracting the negative sentiment scores from the positive ones. Each review was given a unique index for identification.

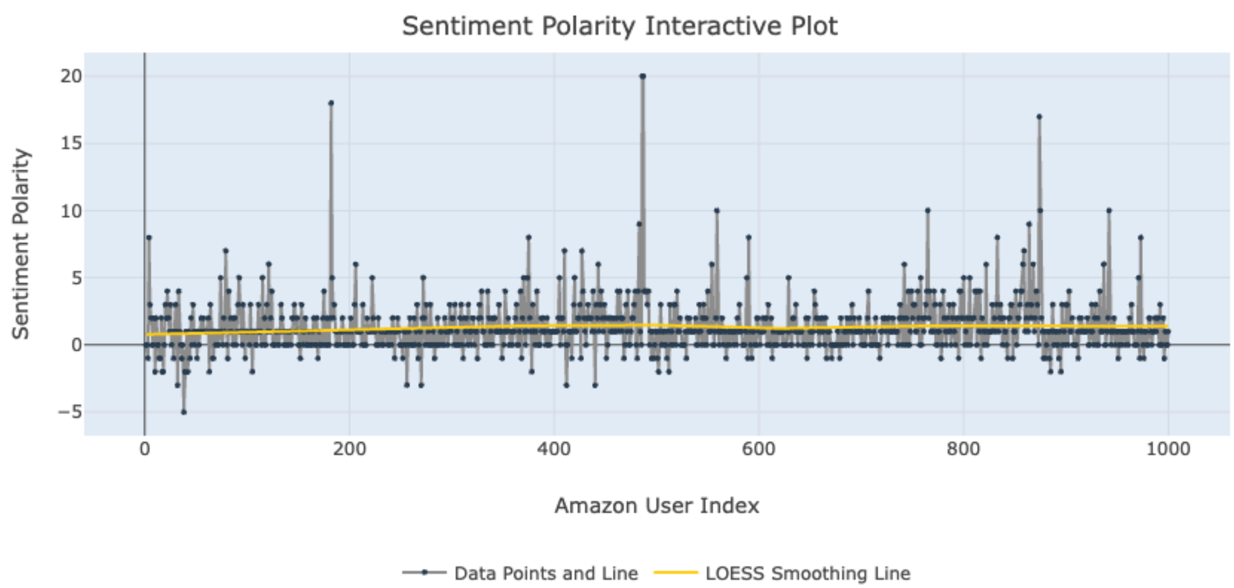

Figure 2: Overall Polarity Plot Filtered by Numbers of Reviews

To represent this data visually, a scatter plot was made where each point indicated a review's sentiment polarity. A LOESS smoothing line was added to highlight the overall trend in sentiment polarity across the dataset. Additionally, the plot was converted into a readable and interactive format, enabling users to zoom in and hover over data points to explore individual reviews in more detail. This interactivity was enhanced with tooltips showing the index and polarity values, along with a filter section that lets users choose how many reviews they want to view.

The distribution of sentiment polarity suggests that most reviews hover around a neutral sentiment, with noticeable outliers showing extreme sentiments. These outliers can be further investigated to understand the reasons behind such strong opinions. The overall trend, indicated by the smoothing line, helps in understanding how sentiments have shifted and whether they have become more positive or negative over time. This trend analysis can provide valuable insights into user behavior and preferences, which are crucial for businesses aiming to improve their products or services.

In summary, the overall polarity plot is an effective tool for visualizing and understanding sentiment in large datasets. It enables the identification of patterns and trends that might not be immediately evident from raw data, offering a deeper understanding of the underlying sentiments expressed in the reviews. The mean polarity score, being around 1.447804, indicates a slight positive sentiment trend in the overall dataset. This means that users generally tend to express more positive sentiments than negative ones when reviewing products in the “All Beauty” category.

B. Time Series Plots

The time series sentiment analysis on the Amazon reviews dataset was conducted to observe changes in sentiment over time, providing insights into evolving customer opinions about beauty products.

To create the time series plots, sentiment analysis was performed on the review texts using R’s “syuzhet” library with the NRC lexicon, which includes sentiments like anger, anticipation, disgust, fear, joy, sadness, surprise, trust, positive, and negative. Sentiment scores were aggregated by year, and the data was cleaned and formatted for accuracy and consistency.

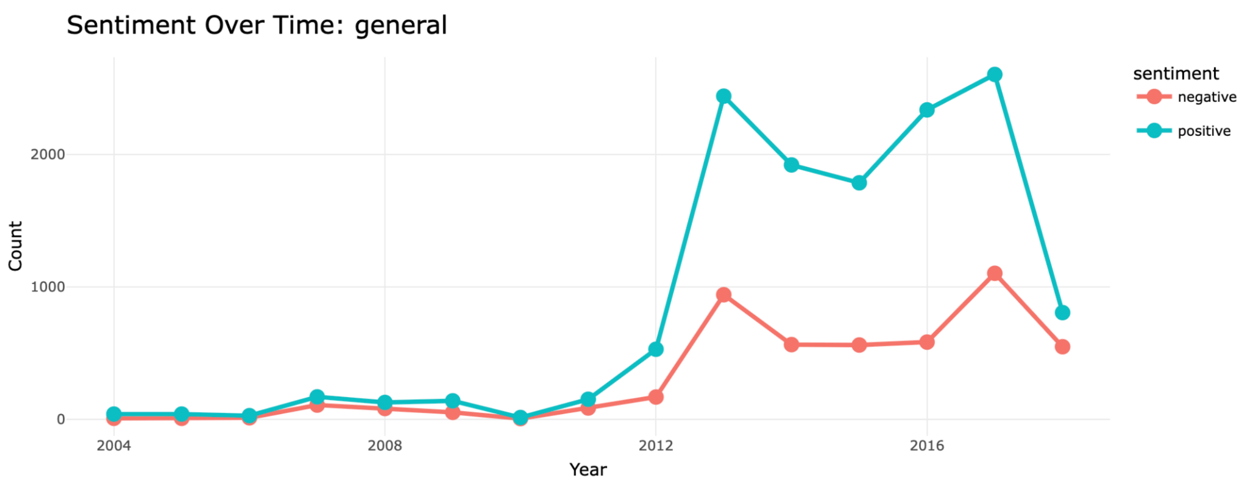

Figure 3: Overall Time Series Sentiment Plot

A line plot was created to visualize this data, where each point represents the aggregated sentiment score for a year. The plot includes both positive and negative sentiments, providing a comprehensive view of sentiment trends over time. The plot is interactive, allowing users to hover over data points for detailed information and zoom in on specific periods. Aesthetics were customized for readability, including color coding and the removal of bold formatting for titles and axis labels.

The general sentiment plot shows a noticeable spike in positive sentiment around 2013, likely due to the release of popular beauty products. Dips in sentiment indicate periods of higher customer dissatisfaction. Sentiment-specific plots reveal significant trends:

1. Disgust, Anger, and Sadness: These negative sentiments show sharp increases around 2012-2013, suggesting that certain products or events triggered strong negative reactions.

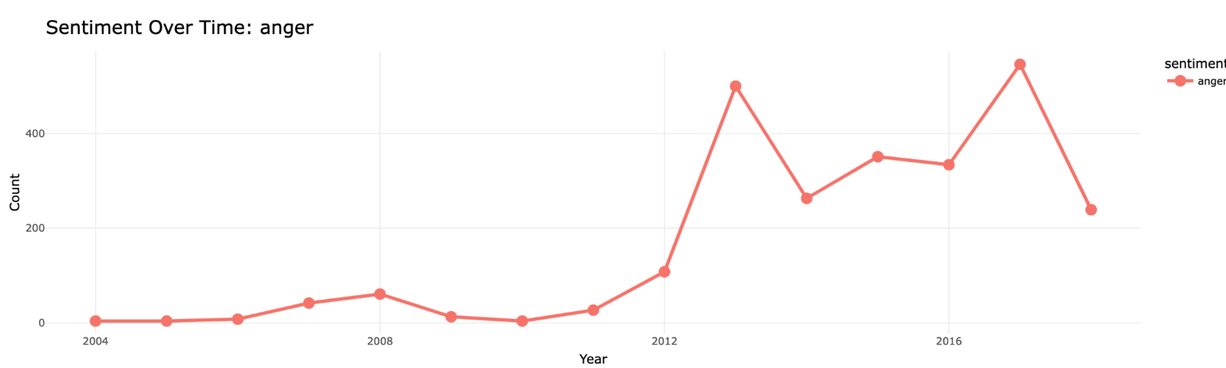

Figure 4: Time Series Plot For Anger

2. Anticipation and Trust: These sentiments peak around 2013 and display a consistent upward trend, indicating growing customer anticipation and trust, possibly due to positive experiences with certain products or brands.

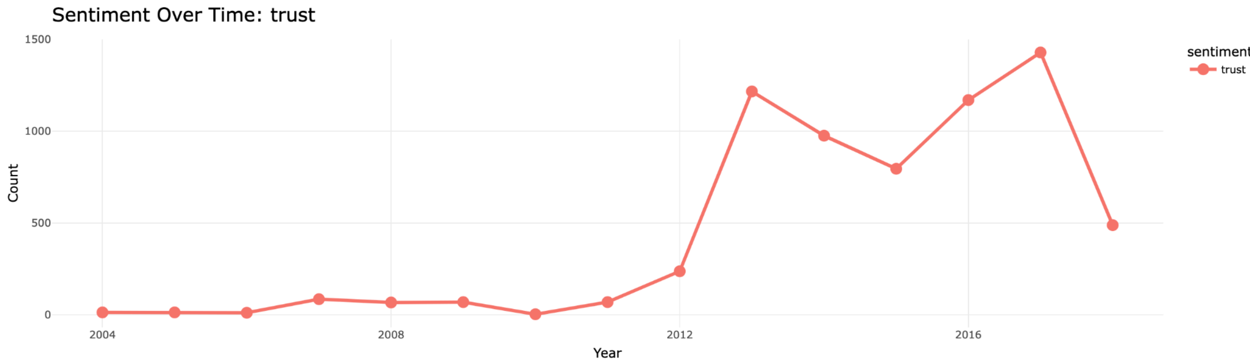

Figure 5: Time Series Plot For Trust

3. Joy: This sentiment shows the most substantial increase around 2012-2013, indicating many users had highly positive experiences with beauty products.

4. Surprise and Fear: These sentiments display volatility, with significant peaks and troughs, suggesting unexpected outcomes with purchases during certain periods.

The time series sentiment analysis reveals an increasing trend in positive sentiment, suggesting that product improvements and new releases have been well-received. Spikes in negative sentiments, such as anger and sadness, highlight periods of customer dissatisfaction that may require further investigation. Including sentiment-specific plots like joy, anger, and anticipation provides deeper insights into customer reactions. Analyzing these trends helps businesses understand customer preferences and areas needing improvement, guiding product development and marketing strategies.

C. Word Clouds (Positive vs. Negative Sentiment)

To create positive and negative sentiment word clouds for Amazon Beauty Reviews, we used trigrams to process and analyze the data. The positive word clouds cover the years from

2007 to 2018, with the exception of 2010. The negative word clouds span from 2011 to 2018. The absence of certain years might be due to the presence of neutral or non-polarizing reviews during those periods, which did not contribute significantly to either positive or negative sentiment.

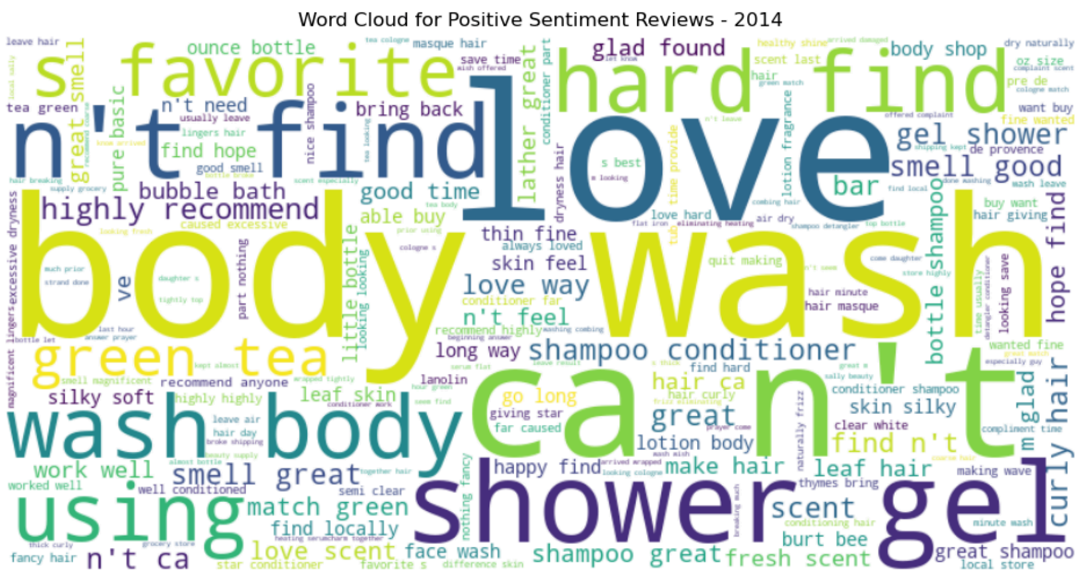

Figure 6: Word Cloud for Positive Sentiment Reviews (2014)

For the positive review word clouds, the data from 2007-2008 prominently features words related to dental products, such as "teeth amazed," "dental floss," and "waterpik." From 2009 to 2018, the focus shifts to bath products, with frequent mentions of terms such as "fragrance," "body wash," and "pampered." The positive word cloud for 2014 showcases a great example of this, as there are several words related to bath products that highlight their luxurious qualities. Therefore, the words in the positive word clouds showcase the customers' positive experiences with both dental and bath products.

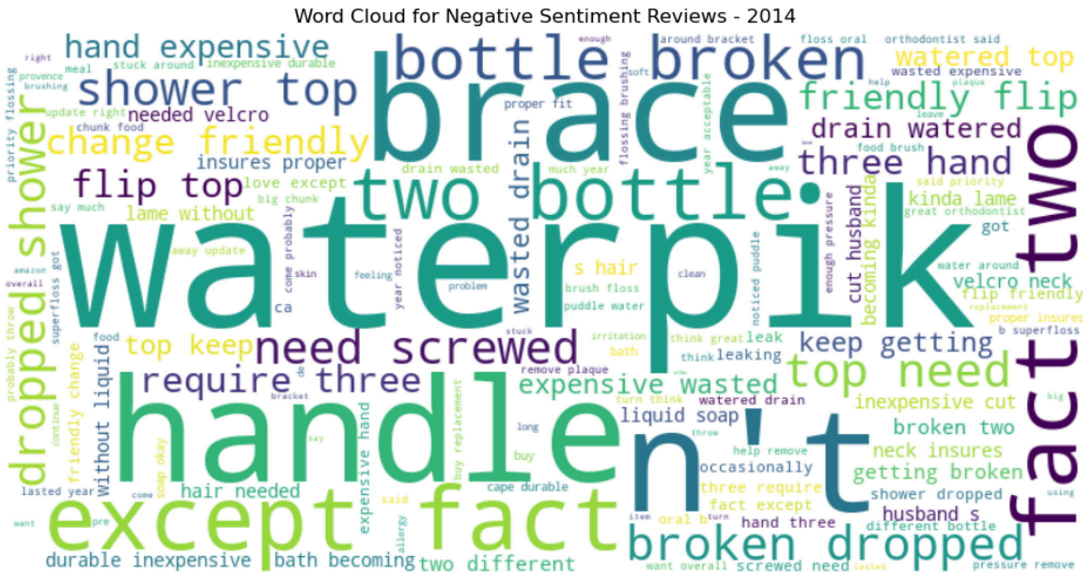

Figure 7: Word Cloud for Negative Sentiment Reviews (2014)

In the negative review word clouds, spanning from 2011 to 2018, many of the most prominent words are associated with body and hair products. Terms such as "stopped hair," "prescribed bad," "bottle broken," "frizz," "greasy," and "hair brittle" appear frequently, indicating common complaints among reviewers. Interestingly, the word "waterpik" appears prominently in negative reviews in 2014, which is noteworthy given that "waterpik" was a positive term in 2007. This evident shift in sentiment towards the “waterpik” could reflect changes in product quality or consumer expectations over time.

D. Word Clouds (Verified vs. Unverified Reviews)

One main characteristic of our dataset is whether or not the review is verified or not. To classify reviewers as “verified” or “unverified,” Amazon checks whether the customer indeed bought the item from Amazon and verify if the customer paid a price that is commonly available to the majority of Amazon customers.[2] From this, we can build two separate word clouds for each year: one from all the verified reviews and the other from all the unverified reviews using trigrams. These word clouds can give us insights of the sentiments of the reviewers as well as common trends that may appear from year to year. While a verified review can give us some reliability of the authenticity of the feedback, it's beneficial to also examine the unverified word clouds as that may give us some important insights as well.

As we have word clouds for data in years 2007-2009 and 2011-2018, we have a total of 11 years and 2 word clouds for each year, resulting in 22 word clouds overall. The word cloud for 2010 is not present due to missing data. Our dashboard includes all of these word clouds but we will analyze a couple below to demonstrate how they are interpreted and its efficiency in helping us with sentiment analysis.

Figure 8: Word Cloud for Unverified Reviews (2018)

Above is the word cloud for unverified reviews in the year of 2018. From the Amazon Beauty reviews of this year, we can see that there are a lot of keywords that relate to hair care such as "soap," "scent," "dandruff," "shampoo," "smell," and "hair" which gives us insights into what many people were purchasing or looking for. Additionally, it appears that there are quite a few positive words such as "great," "love," and "nice" which point to many individuals having a positive sentiment towards these beauty products. However, it is important to keep in mind while examining this word cloud that these are the unverified reviews which means that the authenticity could be questionable and not super reliable.

Figure 9: Word Cloud for Verified Reviews (2018)

This second word cloud corresponds to verified reviews in the year of 2018. Similar to the unverified ones, we see a lot of words relating to hair care such as “smell,” “scalp,” “hair,” “lather,” and “dandruff.” Additionally, we also see quite a few words relating to skin care such as “skin,” “body wash,” “clean.” While the unverified reviews had more of a harmonious positive sentiment, the verified review word cloud doesn’t appear as unified; there are positive words like “good,” “great,” and “good lather,” but there are also negative phrases such as “doesn’t help,” “issue,” and “poor quality.” Overall, it appears that the positive words are appearing more often but there are still some issues that reviews have regarding these verified products. This word cloud is more dependable as it is built off of verified reviews. Also, we can say that for the year of 2018, there are many reviews that are related to skin and hair care with an overall positive sentiment but some negative reviews that are noticeable as well.

Section V: Limitations & Future Work

Our project includes several limitations. Firstly, our sentiment analysis was only limited to beauty products, so we did not get a full understanding of sentiment across all product categories on Amazon. Additionally, slang terms and ambiguous reviews may have caused some misclassifications, leading to potential inaccuracies in our analysis. Another limitation was insufficient detail in the original dataset documentation; since it did not explain product IDs or names, this hindered our ability to do product-specific analysis. Furthermore, we also did not have access to product launch dates and product discontinuation dates, further limiting the scope of our analysis.

If we had more time in the future, we could strengthen our methodology by implementing a BERT model, a deep learning Natural Language Processing model. BERT stands for "Bidirectional Encoder Representations from Transformers." We would train the BERT model to learn words bidirectionally and use it to predict customer sentiment for future years. Expanding our focus to include other product categories, such as electronics, clothing, and furniture, would broaden the scope of our project. Additionally, with access to product names or IDs, we could add more advanced filtering features on our dashboard.

Section VI: References

1. Ni, J. (n.d.). *Amazon Review Data (2018)*. Amazon Review Data. https://nijianmo.github.io/amazon/
2. Amazon verified purchase reviews - Amazon Customer Service. (n.d.). https://www.amazon.com/gp/help/customer/display.html?nodeId=G75XTB7MBMBTXP6W

Section VII: Appendix

This is the link to our GitHub repository, which contains all project files:
https://github.com/Danii-W/Amazon_Beauty_Sentiment_Analysis/tree/main

This is the link to our interactive dashboard: https://mas418final.shinyapps.io/418_shiny/

Here are some screenshots of our dashboard: