Sentiment and Topic Analysis of Sephora Skincare Reviews

A Text as Data analysis using lexicon-based sentiment scoring and unsupervised topic modeling

Amber Ni

2025-05-04

Contents

1 Introduction

1.1 Background

In 2023, 32% of global beauty and personal care sales were made online, up from 21% in 2019 (Statista 2024a). As online shopping continues to gain popularity, beauty companies and brands are increasingly leveraging data analytics to better understand their consumers, develop tailored marketing strategies, and drive innovation. In fact, over 65% of beauty brands reported using AI and data analytics to personalize customer experiences in 2023 (Forbes 2023). With 82% seek out customer reviews before purchasing a beauty or grooming product (Space48 2023)] and 97% of consumers stating that online reviews influence their beauty product purchases (PowerReviews 2023), it has become more important than ever for beauty marketers to align their communications and marketing strategies with consumer preferences and mindsets.

Business analysts are utilizing consumer activity and consumption data to draw inferences and perform predictive analysis, particularly in three key areas: (1) market trends, (2) consumer preferences, and (3) online engagement.

1. Market Trend

Leveraging robust datasets available online across different brands and types of products now enables brands to anticipate emerging beauty trends in the market, giving them a more holistic picture of leading products with further analysis on what is the newest driving force of consumption and allowing them to proactively develop products that meet evolving consumer desires.

2. Consumer preferences

By analyzing consumer purchase history, beauty brands can gain deep insights into what drives purchasing decisions. This includes understanding preferences for product ingredients, pricing, and brand loyalty. Additionally, tracking demographic and geographic preferences enables brands to tailor their strategies for different consumer segments, ensuring that their products align with the diverse needs and expectations of their target audience.

3. Online engagement

User interactions—likes, shares, comments, and reviews—serve as key indicators of brand performance and consumer sentiment. By analyzing engagement metrics across social media and e-commerce platforms, beauty brands can track trending products and identify potential pain points. Additionally, sentiment analysis and recommendation algorithms help brands address concerns, optimize product exposure, and foster deeper consumer interactions.

1.2 Overview of Sephora Online Shop

Sephora's online store serves as a premier platform for personal care and beauty products, offering a diverse and extensive choices from both global luxury brands and emerging labels. It offers over 340 brands and 45,000 products, including skincare, makeup, haircare, and fragrance. Sephora attracts over 50 million unique visitors to its website each month globally (Statista 2024b). With millions of active consumers regularly sharing their experiences through detailed reviews, Sephora provides a rich source of user-generated content that captures authentic customer sentiments and preferences.

The platform's vast selection of skincare products across various categories ensures a comprehensive representation of consumer behavior, making it an ideal dataset for analysis. This combination of high engagement, product variety, and brand diversity positions Sephora's online shop as a valuable resource for data-driven insights into skincare trends, customer satisfaction drivers, and market dynamics.

1.3 Motivation

Customer reviews represent a rich and authentic avenue for understanding consumer behavior, preferences, and expectations, especially in industries like skincare, where personal experiences and subjective perceptions play a critical role in purchasing decisions. Unlike traditional surveys or sales data, reviews provide unsolicited, detailed feedback that captures both the emotional tone and practical considerations influencing consumer satisfaction.

Motivated by the potential of this data, I sought to explore how data science techniques could transform these narratives into actionable insights. Leveraging Sephora's extensive repository of user reviews, this project aims to decode the factors that drive positive sentiment, reveal emerging skincare trends, and understand how elements such as product characteristics, personal skin concerns, and social media influence shape consumer opinions. By doing so, I hope to demonstrate how review analysis benefits brands in optimizing their product development, marketing strategies and consumer support.

1.4 Relevant Previous Work

Recent analyses of Sephora's online reviews primarily focus on understanding customer sentiment, product preferences, and areas for service improvement through natural language processing and machine learning techniques. Projects on platforms like GitHub and Kaggle have employed models such as DistilBERT and PyTorch-based classifiers to perform sentiment analysis on large volumes of review data. These studies aim to extract insights into customer satisfaction, emotional tone, and common pain points, offering Sephora a deeper qualitative view beyond numerical ratings. Articles and customer feedback reports on market research platforms (e.g., from Kimola) also highlight themes like delivery delays, customer service frustrations, and the success of in-store experiences versus online shopping

Overall, the existing body of work highlights a growing interest in using text analytics to enhance Sephora's product strategy and digital customer experience. Unlike numeric ratings which do offer a quick summary of product evaluation, textual reviews offer richer, more nuanced insights by capturing human sentiment and context, providing a richer understanding of customer perceptions and behavior.

2 Research Question

As such, this paper tries to understand **how specific review characteristics shape the emotional tone of skincare product reviews on online beauty platforms like Sephora**. It is expected that elements such as review length, ratings, certain themes, and characteristics of reviewers will significantly influence sentiment, whether positive or negative. The goal is to uncover patterns in emotional expression that can inform product development, marketing strategies, and customer engagement.

3 Data and Method

3.1 Data Source

The dataset used in this project consists of user-generated skincare product reviews scraped from Sephora.com and made available on Kaggle in March 2023. It contains over 60,000 observations, of which a random sample of 10,000 reviews was selected for analysis. This dataset provides both structured attributes (ratings, skin type of the reviewer, product price, brand name) and unstructured text (review texts, review titles).

Variable Category	Variable Name	Symbol	Variable Definition	Unit	Variable Type
Dependent Variable	Normalized Sentiment Score	`norm_sentiment`	Sentiment score of the review, computed using AFINN lexicon and normalized by the length of the review.	Score	Continuous
Explanatory Variables	Token Count	`token_count`	Number of tokens (words) in the review, representing review length.	Token	Continuous
	Product Price	`price_usd`	Listed price of the product.	USD	Continuous
	Skin Type	`skin_type`	Categorical variable indicating the reviewer's skin type (e.g., dry, normal, oily).	N/A	Categorical
	Dominant Topic	`dominant_topic`	Categorical variable indicating the primary topic of the review, identified by topic modeling.	N/A	Categorical
	Mentions Social Media	`mentions_social_media`	Dummy variable. 1 if the review mentions social media, 0 otherwise.	N/A	Binary (Dummy)

Table 1: Description of Variables used in the Study

3.2 Unit of Analysis

The unit of analysis in this project is the individual product review. Each observation represents a single customer's written feedback on a skincare product listed on Sephora's online platform.

3.3 Variables Used

The primary variable of interest is the sentiment score expressed in each review, which captures the overall polarity and strength of sentiment (positive or negative) within the review text. This variable is used both descriptively and as the dependent variable in regression analysis to examine what features influence sentiment expression.

The explanatory variables include a set of review- and reviewer-level characteristics that may influence sentiment expression. These consist of structural features such as review length (token count) and product price, categorical attributes like skin type and dominant review topic, and contextual factors such as whether the review mentions social media. These variables are used to examine how different aspects of review content and context relate to variations in emotional tone (Table 1).

3.4 Methodology

The methodological steps include:

- Descriptive Analysis: Used to explore baseline patterns across brand and skin type groups, helping to contextualize the variation in user feedback before applying formal models.
- Sentiment Analysis: Implemented using the AFINN lexicon, which provides a pre-scored dictionary of English words with sentiment polarity. AFINN was chosen for its simplicity, numeric scoring system (from -5 to $+5$), and effectiveness in capturing emotional tone in consumer-generated content. Sentiment scores were normalized by review length to account for differences in verbosity across reviews.
- Topic Modeling: Latent Dirichlet Allocation (LDA) was used as an unsupervised learning approach to discover latent themes in reviews without requiring pre-labeled data. This method was well-suited for identifying recurring topics in large volumes of unstructured text, enabling the assignment of a dominant topic to each review for further analysis.
- TF-IDF and Log-Ratio Analysis: Used to identify the most distinctive vocabulary across sentiment categories. TF-IDF helped highlight frequently used but distinctive terms, while log-ratio analysis provided a direct comparison of term usage between positive and non-positive reviews.

Table 2: Top 5 Brands with Highest Proportion of Positive Review

brand_name	review_type	count	total_reviews	proportion
REN Clean Skincare	Positive	61	65	0.9384615
Evian	Positive	26	28	0.9285714
Sol de Janeiro	Positive	23	25	0.9200000
Dr. Dennis Gross Skincare	Positive	166	183	0.9071038
bareMinerals	Positive	56	62	0.9032258

Table 3: Top 5 Brands with Highest Proportion of Non-Positive Review

brand_name	review_type	count	total_reviews	proportion
Dr. Jart+	Non-positive	61	204	0.2990196
Isle of Paradise	Non-positive	14	48	0.2916667
The INKEY List	Non-positive	51	181	0.2817680
Skinfix	Non-positive	24	88	0.2727273
Supergoop!	Non-positive	42	160	0.2625000

- Linear Regression: Chosen for its interpretability and ability to quantify the relationship between multiple explanatory variables and the dependent variable. This method allows for assessing the marginal effect of each review feature, while controlling for other variables.

3.5 Limitations to the Dataset/Measures

The dataset is skewed, with over 80% of reviews being positive, which may limit the ability to detect factors influencing negative sentiment. It only covers a single year of data, lacking longitudinal insights or trends over time. In terms of measures, temporal effects such as product seasonality or marketing campaigns for a specific product are not accounted for. Additionally, the sentiment measure is based on a lexicon approach (AFINN), which may not fully capture context, sarcasm, or nuanced consumer expressions. Topic modeling is also constrained by the nature of the data: many reviews are short and contain overlapping vocabulary, making it difficult to extract highly distinct and interpretable topics. Finally, reviewer characteristics (e.g., demographics, purchase history) are not available, limiting control over individual-level heterogeneity.

4 Results

4.1 Summary of Findings

Results reveal that review length, measured by token count, is significantly and negatively associated with sentiment scores. It suggests that longer reviews tend to express more critical or negative emotions. Among the topic-based variables, reviews categorized under "Texture, Fragrance & Makeup" are significantly more positive compared to the baseline category (Moisturization). In contrast, price, skin type, social media mentions, and other dominant topics such as "Problematic Skin" and "Purchase Experience" show no significant effects on sentiment. These findings suggest that how customers talk about certain topics and the length of their expressions plays a more important role in shaping emotional tone than demographic or price-related factors.

Table 4: Proportion of Positive Reviews by Skin Type

skin_type	review_type	count	total_reviews	proportion
combination	Positive	4067	4918	0.8269622
oily	Positive	887	1080	0.8212963
normal	Positive	933	1141	0.8177038
dry	Positive	1313	1623	0.8089957

4.2 Descriptive Analysis of the Ratings

First, reviews were categorized as "positive" (ratings 4–5) or "non-positive" (ratings 1–3) to explore rating patterns at both the brand and individual levels. At the brand level, the analysis focused on identifying which specific brands received the highest share of positive reviews. At the individual level, it focused on how different reviewer skin types received the highest share of positive reviews.

- Brand-level: Five skincare brands with the highest proportion of positive reviews. Interestingly, lesser-known or niche brands, such as REN Clean Skincare, Evian, and Sol de Janeiro, outperform more mainstream competitors in terms of customer satisfaction, with over 90% of their reviews being positive (Table 2 and Table 3).

- Individual-level: Reviewers with combination skin have the highest proportion of positive reviews, which may suggest their skin may be more adaptable to a variety of skincare products. In contrast, reviewers with dry skin report the lowest positive rating share, potentially because of greater sensitivity or difficulty finding products that meet their needs (Table 4).

4.3 Overview of Text Language

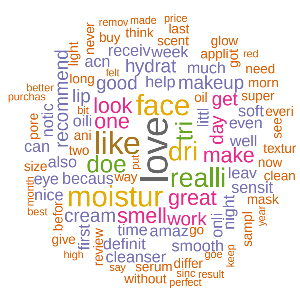

Figure 1: Review Word Cloud

What are these reviews generally about? A word cloud analysis shows that skincare reviewers often express strong positive emotions, using words like "love", "great", and "recommend". Common themes include

Table 5: Top Positive Terms by TF-IDF Weight

	positive_tfidf
use	0.2459646
skin	0.2458082
product	0.2382643
love	0.2176994
feel	0.2126034
moistur	0.2070087
face	0.2014679
like	0.1780894
veri	0.1776177
dri	0.1759039

Table 6: Top Non-Positive Terms by TF-IDF Weight

	nonpositive_tfidf
like	0.2699301
product	0.2624467
just	0.2179181
use	0.2070530
skin	0.2029094
face	0.2027030
dri	0.1978924
doe	0.1948978
realli	0.1922062
tri	0.1889408

hydration and smoothness outcomes ("moisture", "hydration", "glow"), usage routines ("day", "night", "week"), and product types ("cream", "serum", "mask"). This reflects a focus on both emotional satisfaction and skincare results (Figure 1).

4.4 Differences in Vocabulary Between Positive and Non-Positive Reviews

- TF-IDF: What are the most important and distinctive words in positive and non-positive reviews?

TF-IDF analysis shows distinct language patterns by sentiment. Many terms are generic skincare words (e.g., "skin", "face", "product", "use"), which is expected due to TF-IDF being influenced by frequency within the category. Still, positive reviews stand out with emotionally expressive terms like "love", "feel", and "moisture". Non-positive reviews, on the other hand, highlight disappointment or frustration using words like "just", "doe", and "tri", often referring to phrases that reflect unmet expectations (Table 5 and Table 6).

- Log-Ratio of Term Probabilities: How much more (or less) frequent a word is in positive reviews vs. non-positive reviews?

Positive reviews often highlight desirable skincare outcomes like "radiant", "clearer", or "healthy", and reflect feelings of luxury or self-care through terms like "spa" and "treat". In contrast, non-positive reviews focus on dissatisfaction or product failure, using harsh descriptors such as "refund", "trash", "burnt", and "splotch" (Figure 2).

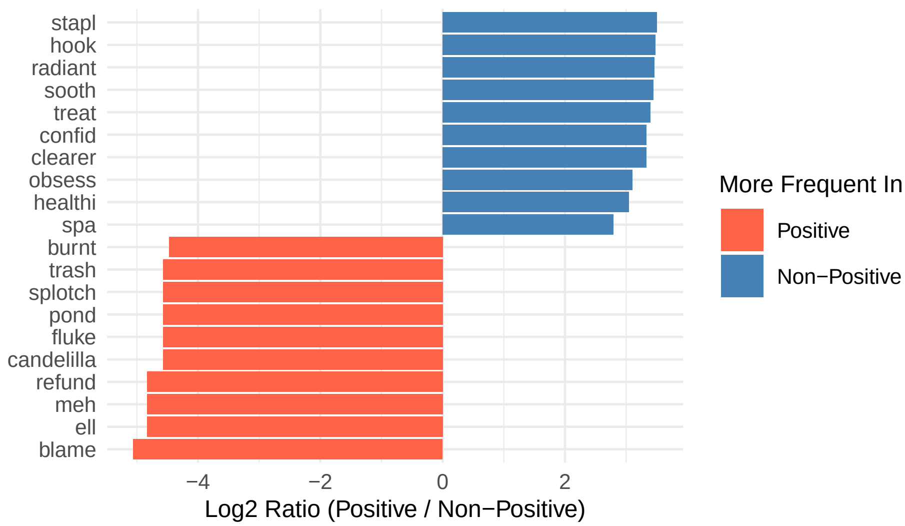

Figure 2: Top 10 Differentiating Terms by Log-Ratio: Positive vs. Non-Positive

4.5 Sentiment Analysis

The AFINN dictionary was used to assign sentiment scores to words on a scale from –5 (most negative) to +5 (most positive). Each review was scored based on the sum of matched terms, and then normalized by the total word count to account for differences in review length.

I then compared the distributions of normalized sentiment scores and star ratings across reviews. Both variables display similarly skewed patterns, with a strong concentration on the positive end—ratings cluster around 4 and 5 stars, while sentiment scores center above zero. This alignment confirms that most reviews are favorable, regardless of whether measured by numeric rating or lexicon-based sentiment scoring (Figure 3).

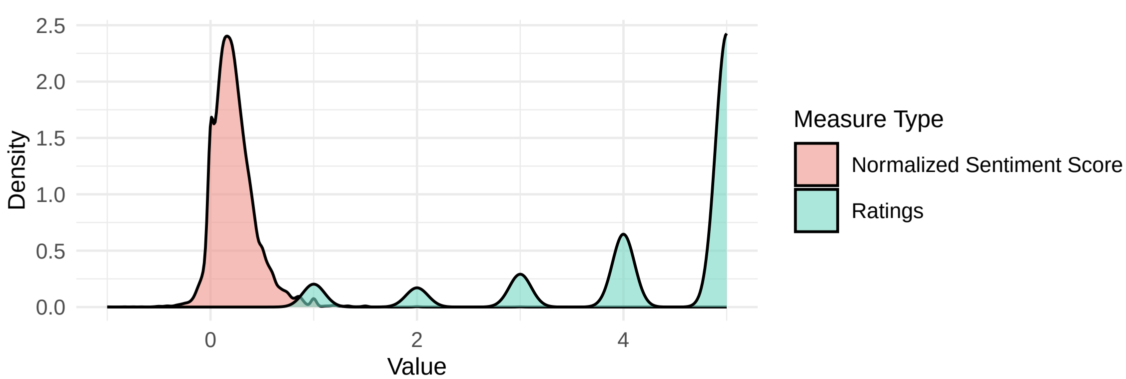

Figure 3: Distribution of Review Ratings and Normalized Sentiment Scores

4.6 Topic Modeling

To examine how review content themes relate to consumer evaluations, I descriptively analyzed both average ratings and sentiment scores across topics generated by LDA. Both in terms of average ratings and sentiment scores, “Texture, Fragrance & Makeup” received the highest evaluations, while “Purchase Experience & Satisfaction” consistently ranked the lowest (Figure 4).

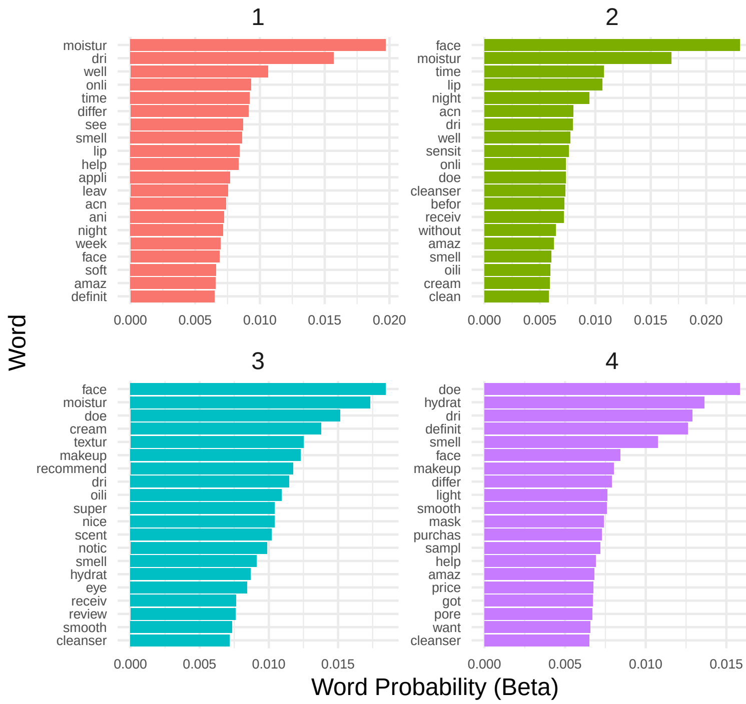

Figure 4: Top Words in Each Topic

- Topic 1: Moisturization - words like “moisture” and “dri”
- Topic 2: Problematic Skin (Acne/Sensitivity) - words like “acn” and “sensit”
- Topic 3: Texture, Fragrance & Makeup - words like “cream”, “textur”, and “makeup”
- Topic 4: Purchase Experience & Satisfaction - words “purchas”, “sampl”, “price”, “order”

4.7 Linear Regression

In this section, I explored the factors associated with sentiment score in skincare product reviews by running a linear regression analysis. The dependent variable was the normalized sentiment score of each review. The main independent variable was the review’s most dominant topic, as identified through LDA topic modeling. I also included several control variables to account for other influences on sentiment: product price, reviewer skin type, review length, and whether the review referenced social media platforms or influencers (e.g., TikTok, YouTube).

The analysis found that longer reviews tend to be associated with lower ratings, while reviews focused on “Texture, Fragrance & Makeup” are significantly linked to higher ratings, which aligns with our descriptive findings previously. In contrast, product price, reviewer skin type, and mentions of social media showed no significant impact on review ratings (Table 7).

Variable Group	Category	Estimate	Std. Error	t value	P value
Intercept	—	0.2937	0.0062	47.43	0.000***
Review Length	token_count	-0.00254	0.00010	-24.98	0.000***
Price	price_usd	7.69e-06	6.01e-05	0.13	0.898
Skin Type	dry	0.00108	0.00576	0.19	0.852
	normal	-3.80e-06	0.00661	-0.00	1.000
	oily	-0.01001	0.00678	-1.48	0.140
	(baseline: combination)				
Dominant Topic Category	Problematic Skin (Acne/Sensitivity)	0.00017	0.00619	0.03	0.978
	Purchase Experience & Satisfaction	-0.00315	0.00609	-0.52	0.605
	Texture, Fragrance & Makeup	0.05848	0.00599	9.77	0.000***
	(baseline: Moisturization)				
Social Media Mention	Yes (Dummy)	0.00981	0.02225	0.44	0.659

Table 7: Regression Results: Review Characteristics and Sentiment

5 Discussion

Our analysis reveals that Texture, Fragrance & Makeup consistently received the highest average sentiment and ratings across skincare review topics. Reviews mentioning social media showed no statistically significant sentiment difference, while token count (i.e., review length) was a strong negative predictor of normalized sentiment score. Customers with dry skin were more likely to leave non-positive reviews, suggesting unmet needs in hydration. Interestingly, smaller or niche skincare brands such as REN Clean Skincare and Sol de Janeiro outperformed many well-known brands in terms of positive review proportions.

This project contributes a scalable and interpretable NLP framework for extracting structured insights from unstructured consumer reviews.

Our findings offer actionable guidance for product and marketing teams:

- The clear dissatisfaction from dry skin customers suggests an opportunity for brands to improve hydration-oriented product lines and tailor messaging to address this group's unmet needs.
- The strong performance of lesser-known brands challenges assumptions about market dominance, indicating that consumer trust and satisfaction may hinge more on efficacy than name recognition.
- Despite expectations, price did not significantly affect review sentiment, which implies that customers care more about value and results than price point alone. This insight could inform pricing strategy and communication emphasis.
- Lastly, token count's negative association with sentiment may reflect that longer reviews are more likely to express complaints—useful for customer service teams monitoring detailed feedback.

Future work could incorporate time trends to capture evolving skincare preferences and influencer effects over time. Expanding the analysis to multilingual reviews or comparing cross-platform sentiment (e.g., Sephora vs. Ulta) may also enhance the generalizability of the findings.

6 All Packages used

This analysis was conducted using a variety of R packages that supported different stages of the workflow. Data manipulation and cleaning were primarily handled using the tidyverse (Wickham 2023b), including dplyr (Wickham et al. 2023) and stringr (Wickham 2023a). For text analysis, the quanteda ecosystem was extensively used, including quanteda (Benoit, Watanabe, Wang, Nulty, et al. 2025), quanteda.corpora (Benoit 2020), quanteda.textstats (Benoit, Watanabe, Wang, Lua, et al. 2024), quanteda.textmodels (Benoit, Watanabe, Wang, Perry, et al. 2025), and quanteda.textplots (Benoit, Watanabe, Wang, Obeng, et al.

2024). Topic modeling was performed using topicmodels (Grün and Hornik 2024), and preprocessing steps were supported by textclean (Rinker 2018) and tidytext (Robinson and Silge 2024). The textdata package (Hvitfeldt 2024) was used to load lexicon dictionaries.

Reference

Benoit, Kenneth. 2020. *Quanteda.corpora: A Collection of Corpora for Quanteda.* http://github.com/quanteda/quanteda.corpora.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Jiong Wei Lua, and Jouni Kuha. 2024. *Quanteda.textstats: Textual Statistics for the Quantitative Analysis of Textual Data.* https://quanteda.io.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, Akitaka Matsuo, and William Lowe. 2025. *Quanteda: Quantitative Analysis of Textual Data.* https://quanteda.io.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2024. *Quanteda.textplots: Plots for the Quantitative Analysis of Textual Data.* https://CRAN.R-project.org/package=quanteda.textplots.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Patrick O. Perry, Benjamin Lauderdale, Johannes Gruber, and William Lowe. 2025. *Quanteda.textmodels: Scaling Models and Classifiers for Textual Data.* https://github.com/quanteda/quanteda.textmodels.

Forbes. 2023. "How AI Is Transforming the Beauty Industry." 2023. https://www.forbes.com/councils/forbestechcouncil/2023/11/17/advanced-ai-the-next-frontier-in-beauty-technology/.

Grün, Bettina, and Kurt Hornik. 2024. *Topicmodels: Topic Models.* https://CRAN.R-project.org/package=topicmodels.

Hvitfeldt, Emil. 2024. *Textdata: Download and Load Various Text Datasets.* https://emilhvitfeldt.github.io/textdata/.

PowerReviews. 2023. "Beauty Shopper Study." https://www.powerreviews.com/consumers-depend-on-reviews/.

Rinker, Tyler. 2018. *Textclean: Text Cleaning Tools.* https://github.com/trinker/textclean.

Robinson, David, and Julia Silge. 2024. *Tidytext: Text Mining Using Dplyr, Ggplot2, and Other Tidy Tools.* https://juliasilge.github.io/tidytext/.

Space48. 2023. "Beauty, Grooming & Cosmetics Online Shopping Consumer Survey Findings." https://www.space48.com/wp-content/uploads/2023/01/Beauty-Grooming-Cosmetics-Online-Shopping-Consumer-Survey-findings-2023-1.pdf.

Statista. 2024a. "E-Commerce Share of Beauty and Personal Care Sales Worldwide from 2019 to 2023." https://www.statista.com/statistics/1414685/yoy-online-beauty-revenue-growth-worldwide/.

———. 2024b. "Sephora.com Traffic Analytics & Market Share." https://www.statista.com/statistics/1383955/distribution-of-sephora-website-traffic-by-country/#:~:text=As%20of%20April%202024%2C%20the%20United%20States,16%20percent%20of%20the%20total%20traffic%20share.

Wickham, Hadley. 2023a. *Stringr: Simple, Consistent Wrappers for Common String Operations.* https://stringr.tidyverse.org.

———. 2023b. *Tidyverse: Easily Install and Load the Tidyverse.* https://tidyverse.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation.* https://dplyr.tidyverse.org.