How do review characteristics affect the emotional tone in skincare product reviews?

AMBER NI*, Georgetown University

1 INTRODUCTION

1.1 Background

In 2023, 32% of global beauty and personal care sales were made online, up from 21% in 2019 (Statista, 2024). 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, How AI Is Transforming The Beauty Industry, 2023). With 82% seek out customer reviews before purchasing a beauty or grooming product (Space48, Beauty, grooming & cosmetics online shopping consumer survey findings, 2023) and 97% of consumers stating that online reviews influence their beauty product purchases (PowerReviews, Beauty Shopper Study, 2023), it has become more important than ever for beauty marketers to

Author's Contact Information: Amber Ni, xn8@georgetown.edu, Georgetown University.

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.

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. From 2016 to 2024, sephora.com e-commerce net sales grew from 580 million U.S. dollars to around 3.4 billion U.S. dollars in sales (Statista, 2025). Sephora attracts over 50 million unique visitors to its website each month globally (Similarweb, 2024). 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.

1.2 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.

1.3 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.

1.4 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.

1.5 Overview of the Project

This project explores how specific characteristics of skincare product reviews, such as review length, topical content, social media references, and reviewer demographics, affect the emotional tone expressed in the text. Using a subset of 10,000 Sephora reviews collected in March 2023, the analysis begins with a descriptive overview of which brands and reviewer skin types are associated with higher ratings. I then apply dictionary-based sentiment analysis, topic modeling, and regression techniques to evaluate whether these descriptive patterns align with the underlying sentiment and to uncover key factors that shape the emotional tone of customer feedback.

By analyzing the language consumers use, this study sheds light on the deeper emotional undercurrents that influence purchasing behavior and product perception, providing value to both marketers and product designers in the online beauty industry.

2 DATA AND METHOD

2.1 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).

2.2 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.

2.3 Variable of Interest

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.

2.4 Data Wrangling Steps

- A random subset of 10,000 reviews was selected from the full dataset of over 60,000 skincare product reviews.
- The text was preprocessed by removing punctuation, numbers, contractions, and non-ASCII characters, then converted to lowercase and cleaned using stop word removal, stemming, and symbol replacement.

3 ANALYSIS

3.1 Description of Methods and Tools used

The methodological steps include:

- **Descriptive Analysis**: To summarize patterns in ratings by brand and skin type.
- **Sentiment Analysis**: Implemented using the AFINN lexicon, which assigns numeric sentiment scores to words. These scores were normalized by review length to account for variation in text volume.
- **Topic Modeling**: Latent Dirichlet Allocation (LDA) was applied to identify common themes across reviews, with each review assigned a dominant topic label for further analysis.
- **TF-IDF and Log-Ratio Analysis**: Used to identify distinguishing vocabulary between positive and non-positive reviews.
- **Linear Regression**: Conducted to estimate the effect of review characteristics on the normalized sentiment score.

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

brand_name <chr>	review_type <chr>	count <int>	total_reviews <int>	proportion <dbl>
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

Fig. 1. Top 5 Brands with Highest Proportion of Positive Reviews

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.

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:

brand_name <chr>	review_type <chr>	count <int>	total_reviews <int>	proportion <dbl>
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

Fig. 2. Proportion of Positive and Non-Positive Reviews by Skin Type

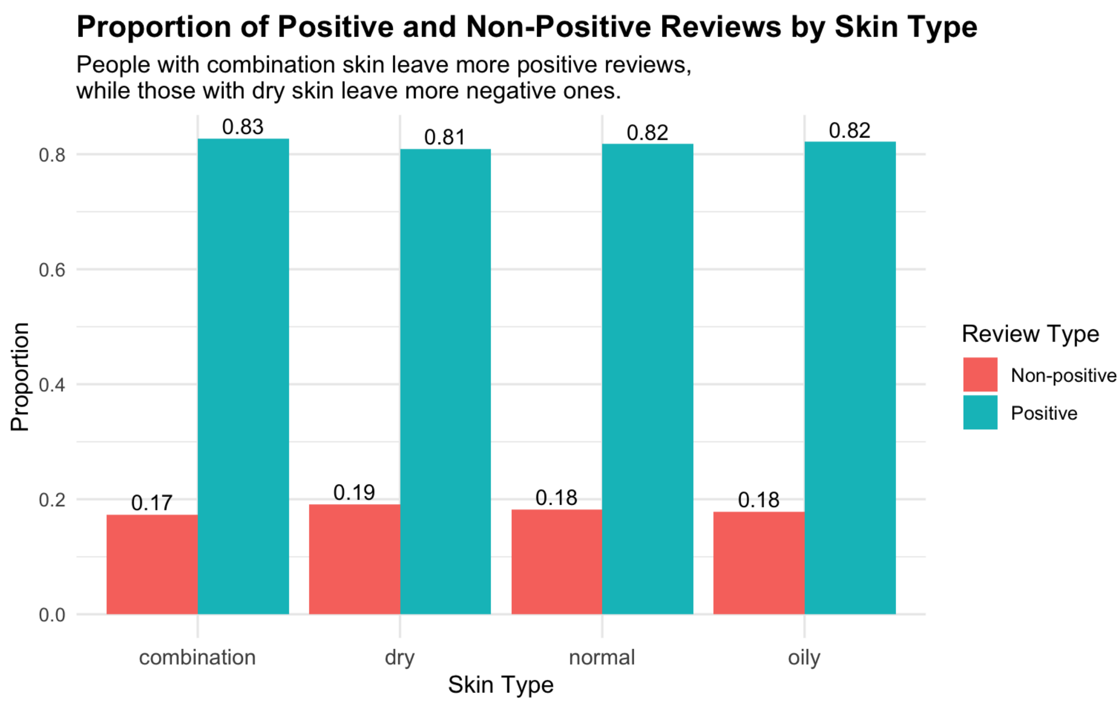

Fig. 3. Review Word Cloud

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 (Fig. 1 and Fig. 2).

- Individual-level:

Fig. 4. Top 5 Brands with Highest Proportion of Non-Positive Reviews

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 (Fig. 4).

4.3 Overview of Text Language

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

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

Fig. 5. Top Positive Terms by TF-IDF Weight

"recommend". Common themes include 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 (Fig. 4).

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

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

Fig. 6. Top Non-Positive Terms by TF-IDF Weight

like “just”, “doe”, and “tri”, often referring to phrases that reflect unmet expectations (Fig. 5 and Fig. 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” (Fig. 7).

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

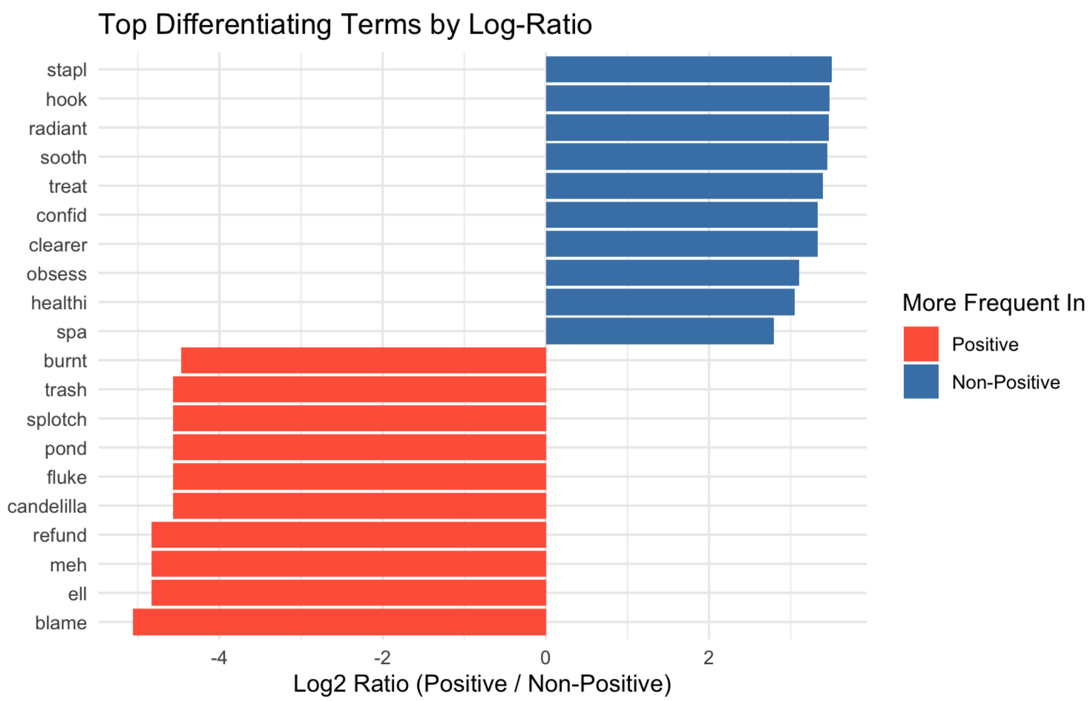

Fig. 7. Top Differentiating Terms by Log-Ratio

matched terms, and then normalized by the total word count to account for differences in review length.

4.6 Topic Modeling

- 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 like "purchase", "sample", "price", and "order"

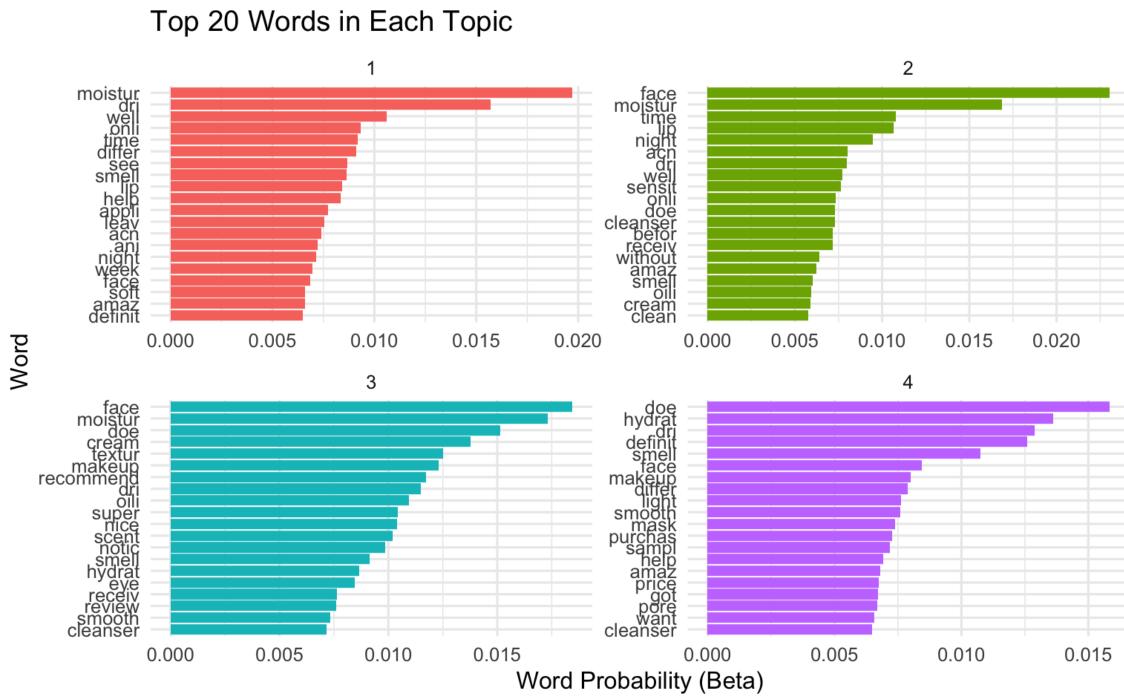

Fig. 8. Top 20 Keywords Across Skincare Review Topics

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 (Fig. 8 and Fig. 9).

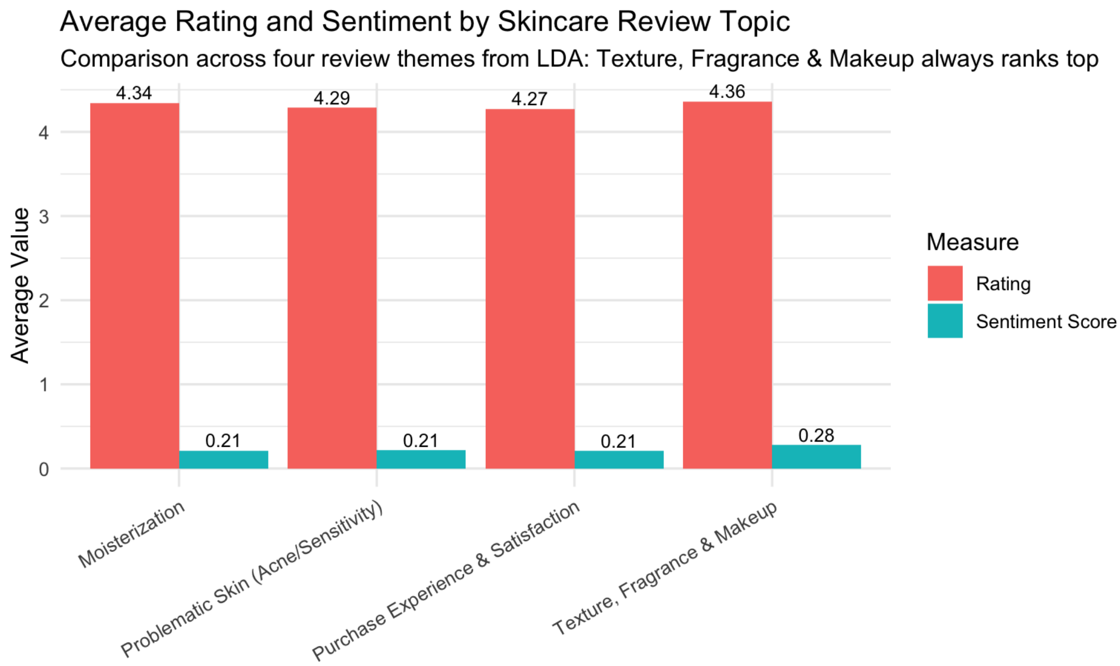

Fig. 9. Average Rating and Sentiment by Skincare Review Topic

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 (Tab. 1).

Table 1. Linear Regression Results

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.2937	0.0062	47.43	$<2e\text{-}16^{***}$
token_count	-0.00254	0.00010	-24.98	$<2e\text{-}16^{***}$
price_usd	7.69e-06	6.01e-05	0.13	0.898
skin_type: dry	0.00108	0.00576	0.19	0.852
skin_type: normal	-3.80e-06	0.00661	-0.00	1.000
skin_type: oily	-0.01001	0.00678	-1.48	0.140
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	$<2e\text{-}16^{***}$
Mentions Social Media	0.00981	0.02225	0.44	0.659

Residual SE = 0.1985, R^2 = 0.08695, Adj. R^2 = 0.08599
F-statistic = 90.31 on 9 and 8535 DF, p-value $< 2.2e{-}16$
Signif. codes: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

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.