The Link Between Customer Reviews and Beauty Awards: An Analysis

Rachel Lewis

Abstract — Every October, Allure Magazine releases a list of its annual awards, given to one top beauty product in every category. This study examines customer sentiment toward Allure Magazine's 2024 Best of Beauty award winners by analyzing Sephora customer reviews. A 2023 survey found that 88% of consumers in the United States rely on ratings and reviews when purchasing beauty products (PowerReviews, 2023a). Research also highlights the strong sales impact of winning an Allure award, such as Fenty Beauty seeing a 132% increase in sales the week after winning a 2018 Best of Beauty award (FasterCapital, 2024). This study compares Allure's Best of Beauty award winners for 2024 with customer sentiment, using Sephora reviews to address three key research questions. How does customer sentiment differ between Best of Beauty award winners and non-award winners? Using customer ratings and sentiment scores, how accurately can we predict which beauty products will be Best of Beauty winners? Using keyword analysis, are there specific keywords in customer reviews that distinguish Best of Beauty winners from non-winners, and can these keywords help predict future award winners? Practical applications of this research include giving brands information to optimize their product marketing strategies based on customer sentiment and potentially best position themselves to win an Allure Best of Beauty award. To answer the research questions, I web scraped over 778,000 records of customer review data from the Sephora website across over 7,000 products, 85 of which were Best of Beauty 2024 award winners. After Exploratory Data Analysis to compare the data sets of award-winning reviews versus non-winners, I used term frequency-inverse document

frequency (TF-IDF) to quantify the importance of terms in Sephora customer review data. I then used Sentiment Analysis (VADER) to analyze how customer sentiment differs between award winners and non-award winners. Finally, I created several classification models - logistic regression, random forest, and XGBoost - to answer how accurately we can predict which products will be Best of Beauty winners based on customer ratings and sentiment scores. This analysis found a statistically significant increase in positive customer sentiment towards award-winning products versus non-winners; however, customer review feedback, comprised of ratings (1-5), review text, and sentiment analysis, is insufficient for predicting Allure Best of Beauty award winners.

1 INTRODUCTION

1.1 Background Information

Reading consumer reviews is now a significant part of shopping, especially in the beauty industry. Fellow customers provide valuable feedback and anecdotes of their experience that help guide potential buyers' decisions. A 2023 survey of over 8,000 U.S. consumers found that more than 90% of respondents consider customer ratings and reviews when making a purchase decision, and 45% of consumers will not purchase a product if no reviews are available (PowerReviews, 2023b). 70% of shoppers in the beauty market use ratings and reviews to learn about and discover products they have not purchased or tried before (PowerReviews, 2023a). A survey of over 26,000 beauty shoppers found ratings and reviews to be the factors consumers trust the most when making a purchasing decision (PowerReviews, 2023a).

Allure Magazine, a leading authority in beauty journalism, is a highly influential source in beauty product evaluation. Each year, Allure announces its Best of Beauty awards, selecting and highlighting top products in various categories. Winning this prestigious award can significantly boost a brand's visibility, credibility, and sales. For instance, after Fenty Beauty's Pro Filt'r Soft Matte Longwear Foundation won an Allure Best of Beauty award in 2018, product sales increased by 132% within a week (FasterCapital, 2024).

Sephora, a global leader in beauty retail with over 500 stores in the United States, is a key player in distributing many Best of Beauty winners (Inside Sephora, n.d.). The 2023 Cosmetify Index revealed that Sephora had the highest organic search traffic and the largest number of Instagram followers among beauty retailers (Cosmetify, n.d.). Of the 7,587 products submitted for the 2024 Best of Beauty awards, Allure selected 357 winners. Of the 357 winning products, 85 are sold at Sephora.

1.2 Research Objectives

This study aims to analyze customer sentiment surrounding the 2024 Best of Beauty award winners by examining Sephora reviews. Specifically, it will address three key research questions to understand how customer opinions compare to Allure's selections.

1. How does customer sentiment differ between Best of Beauty award winners and non-award winners?
2. How accurately can we predict which products will be Best of Beauty winners based on customer ratings and sentiment scores?
3. Are there specific keywords in customer reviews that distinguish Best of Beauty winners from non-winners, and can these keywords help predict future award winners?

1.3 Practical Significance

Having established that winning an Allure Best of Beauty award can positively impact product sales, this analysis will seek to identify how brands can optimize their product marketing strategies based on customer sentiment and potentially best position themselves to win an Allure Best of Beauty award.

This analysis will investigate whether there are common factors within the customer sentiment for Best of Beauty winners that do not exist for non-winners. The insights from this data analysis could inform a brand's marketing strategy based on what customers most commonly reference in customer feedback.

1.4 Study Limitations

This analysis is limited to the award-winning products sold at Sephora, allowing for review, rating, and sentiment data for 85 of the 357 products. A set number of award-winning products limits the analysis to a small pool of customer ratings and review text, compared to a much larger pool of customer reviews available for non-winners, leading to a class imbalance in the classification models.

Additionally, brands are increasingly faking or inflating their product ratings as the beauty industry has become heavily reliant on customer reviews (Burt, 2024). In October 2019, the Federal Trade Commission (FTC) charged Sunday Riley Skincare, a skincare brand sold at Sephora, "with misleading consumers by posting fake reviews of the company's products on a major retailer's website, at the CEO's direction, and by failing to disclose that the reviewers were company employees" (Federal Trade Commission, 2019a). The complaint filed by the FTC detailed several specific instances from November 2015 to April 2018 in which the Sunday Riley CEO and management team instructed employees and interns to post artificial positive reviews of Sunday Reily products to Sephora using fake accounts (Federal Trade Commission, 2019b). According to the FTC, after Sephora removed the fake employee-written reviews, Sunday Riley Skincare obtained a Virtual Private Network to allow employees to hide their IP addresses and location when writing reviews (Federal Trade Commission, 2019b).

In 2020, the FTC settled with Sunday Riley Skincare and prohibited the company from misrepresenting the status of any endorser or person reviewing a product they are selling (Federal Trade Commission, 2020).

Recently, the FTC took regulations a step further, and in August of 2024, announced a "final rule that will combat fake reviews and testimonials by prohibiting their sale or purchase and allow the agency to seek civil penalties against knowing violators" (Federal Trade Commission, 2024). According to the FTC, the final rule, 16 CFR Part 465: Trade Regulation Rule on the Use of Consumer Reviews and Testimonials (2024), prohibits:

- Fake or False Consumer Reviews, Consumer Testimonials, and Celebrity Testimonials
- Buying Positive or Negative Reviews
- Insider Reviews and Consumer Testimonials
- Company-Controlled Review Websites
- Review Suppression

Sephora takes steps to reduce fake reviews, including allowing consumers to indicate whether they received a product for free in exchange for a review and automatically flagging a high volume of reviews for one product coming from the same IP address (Chia, 2019). Despite Sephora's efforts to reduce the number of fake reviews on their website, the potential for fake reviews is a limitation of the study.

1.5 Data and Data Mining Processes

1.5.1 *Data Used in the Study*

This analysis was built around three key tables:

1. Allure Best of Beauty 2024 Winners at Sephora
2. Sephora Product IDs
3. Sephora Customer Review Data

The Allure Best of Beauty 2024 Winners at Sephora table contains data collected manually by searching for all 357 Allure Best of Beauty Winners on Sephora's website and manually gathering the Product IDs from the product URL. This table holds the Category, Brand, Product Name, and Product ID for 85 Allure Best of Beauty 2024 award winners sold at Sephora.

The Sephora Product IDs table contains data collected from Sephora's website using a Python script to collect data from the API: https://www.sephora.com/api/v3/users/profiles/current/product/. The Python script looped numerically in increments of 1, passing each number to the API as a possible Product ID. If the API returned a valid product name for the ID, then the script stored the Product ID in a Snowflake table. If not, the loop continued to the following number. This table holds the Product Name, Brand Name, and Product ID for over 7,700 products sold at Sephora in the same categories as the Allure Best of Beauty award winners.

The Sephora Customer Review Data table data was mined using a Python script to pass the Product IDs stored in the other two tables to the API https://api.bazaarvoice.com/data/reviews.json. The script captured and stored all relevant customer review data in the table for further analysis. This table contains almost 780,000 records of customer review data, including Product ID, Review

Test, Rating (1-5), Recommended (Y/N), Submission Time, Skin Tone, Eye Color, Skin Type, and Hair Color.

While running the models used in the analysis, I created additional tables, including tables to store TF-IDF data for winning and non-winning products. I later used the data in these tables as input in the classification models. I also created a table to store the TF-IDF keywords, which allowed for more straightforward processing in the Python scripts used to build the models.

Tables 1, 2, and 3 below contain data samples of the main three tables used in this analysis.

Category	Brand	Product Name	Product ID
Hair	K18	Damage Shield pH Protective Shampoo	P509691
Body	Kate McLeod	The Pebble Solid Bath & Shower Oil	P508743
Skin	Kate Somerville	KateCeuticals SuperCell Rejuvenation Serum	P510352
Scent	Maison Margiela Fragrances	Replica From the Garden Eau de Toilette	P507949

Table 1- Allure Best of Beauty 2024 Winners at Sephora table sample

Product ID	Brand Name	Product Name
P481969	DIOR	Dior Addict Shine Lipstick
P202633	Anastasia Beverly Hills	Brow Wiz® Ultra-Slim Precision Eyebrow Pencil
P387589	Too Faced	Hangover Replenishing Face Primer

***Table 2* -** Sephora Product IDs table sample

Product ID	Review Text	Rating	Is Recommended (T/F)	Submission Time	Skin Tone	Eye Color	Skin Type	Hair Color
P421998	The store gave me the sample and I used it. I fall in love with it. My skin is very sensitive, but this oil works great on my skin.	5	TRUE	2024-11-03 01:32:43.000	medium	hazel	combination	black
P506273	It's a rich complex smell but doesn't last at all. Also, it's not fresh. It has kind of a chemical smell, like a room spray.	3	FALSE	2024-06-19 04:28:42.000	light	brown	combination	

P506273	I love the smell of this stuff and I really enjoy a different take on a clean scent. But the fragrance had an actual hair IN it. So gross	1	FALSE	2024-07-02 04:48:45.000	fairLight	brown	normal	brown

Table 3 **-** Sephora Customer Review Data table sample

1.5.2 *Data Mining Process*

I compiled all the data used in the analysis in one of three ways. Data were either collected manually using the Sephora and Allure websites, using Python scripts that pulled Sephora website data from public APIs found on the Sephora network, or were an output of a model.

To efficiently collect a substantial amount of customer review data, I wrote Python scripts to scrape review data from the Sephora website. I chose the web scrape method for three reasons:

1. This analysis required very recent customer review data, as Allure released the 2024 Best of Beauty awards at the beginning of October. The publicly available Sephora customer review data sets I found were all generated before 2024.

2. I have written web scrape scripts professionally and am comfortable gathering the data I need using this method.

3. I started by collecting 7,500 Sephora Product IDs. Many products had more than 100 reviews, and manually collecting that amount of data would take an unreasonable amount of time.

1.5.3 *Data Mining Challenges*

I initially attempted to write web scraping scripts using Beautiful Soup before adding Selenium to handle pop-ups. Selenium works well at handling JavaScript on a web page; in this case, pop-ups on the Sephora website, but is not very performant, so I quickly realized this option was not scalable to the size of this project.

To efficiently gather customer review data, I needed a list of Sephora product IDs to feed to the API URL that contained customer review data. Products on the

Sephora website that have won a Best of Beauty award do display a badge, shown below in Figure 1, similar to the Clean at Sephora badge.

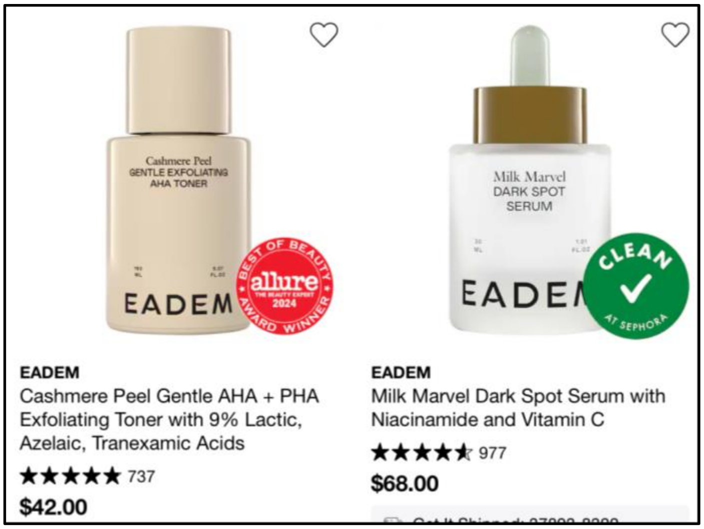

Figure 1 - Allure Best of Beauty 2024 Award Winner badge shown on Sephora's website

To efficiently gather customer review data, I needed a list of Sephora product IDs to feed to the API URL that contained customer review data. Products on the Sephora website that have won a Best of Beauty award display a badge similar to the Clean at Sephora badge, as shown below in Figure 1. Unfortunately, using a web scraper to crawl through every single product page on Sephora's website, looking for product IDs correlating to products with an Allure Best of Beauty badge, was not practical. Compounding matters, Sephora's product IDs are not in perfect numerical order. The product IDs jump around in character count, and the numbers are inconsistent. To gather the product IDs for award-winning products, I manually searched Sephora's website for all winning products and captured the product ID from the product URL.

After looking through Sephora's website, I found an API with a public endpoint containing a JSON file holding all the review data for a particular product. With a list of product IDs, I wrote a Python script to pull the reviews for a singular product (this requires looping through various limits and offsets). The script stores the raw data in a JSON file. I then wrote a second Python script to parse the JSON file and save the pertinent fields for analysis in a CSV format. I slowly built out the script to be more scalable and efficient, sending up to 250 Sephora product IDs to the API at a time. As the data set grew, I moved from storing the data locally to storing the data in Snowflake, increasing my data storage and computing capacity.

The result was a Snowflake table of Sephora customer review data for each product I included in the analysis.

2 PROBLEM STATEMENT

This study analyzes how customer sentiment, as determined by Sephora customer reviews, aligns with the Allure Best of Beauty awards and whether customer sentiment can predict award outcomes. This study analyzes Sephora reviews to explore differences in sentiment between award winners and non-winners, evaluate the predictive accuracy of customer ratings and sentiment scores, and identify distinguishing keywords in reviews that could forecast future award winners.

3 METHODOLOGY

3.1 Exploratory Data Analysis

I started exploratory data analysis by comparing customer ratings for winners and non-winners. The customer review data includes a rating of 1-5, where 1 is the lowest rating a customer can give a product and 5 is the highest. The comparison of customer ratings for winners and non-winners, shown in Figure 2, shows that products that won an Allure Best of Beauty 2024 award have a slightly higher proportion of 5-star reviews and a lower proportion of 1, 2, and 3-star reviews than non-winners.

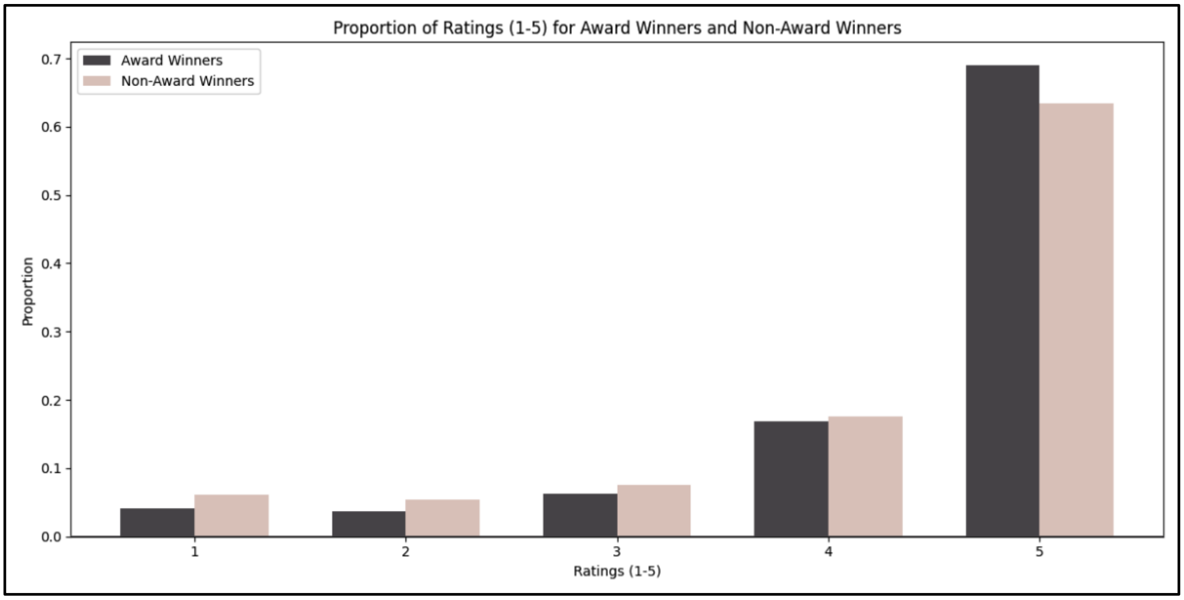

Figure 2 - A comparison of customer ratings (1-5) for winners and non-winners

Next, I compared the review length text between winners and non-winners, as shown in Figure 3. The comparison of the lengths of customer reviews, as calculated by character count, shows that the award winners have a higher percentage of reviews with 50-300 characters, with non-winners having a slightly higher percentage with more than 300 characters.

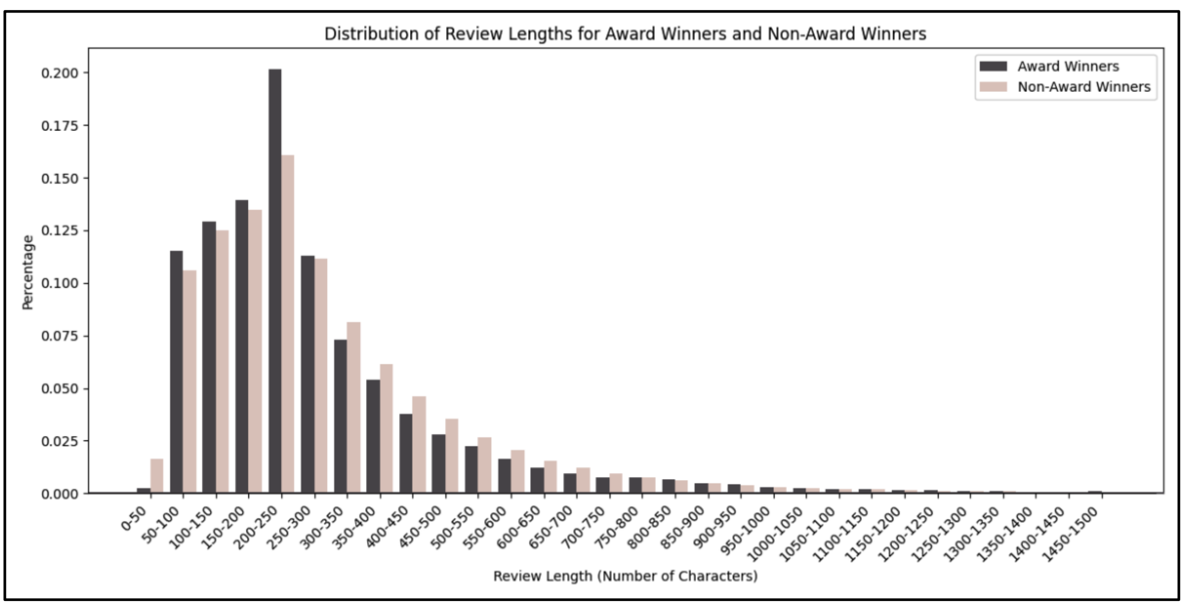

Figure 3 - A comparison of the lengths of customer reviews

I also examined the relationship between length and rating for winners and non-winners. This comparison shows that five and four-star ratings have slightly longer reviews, and products that won an Allure Best of Beauty award also tend to have longer reviews than non-winners. This comparison is shown in Figure 4.

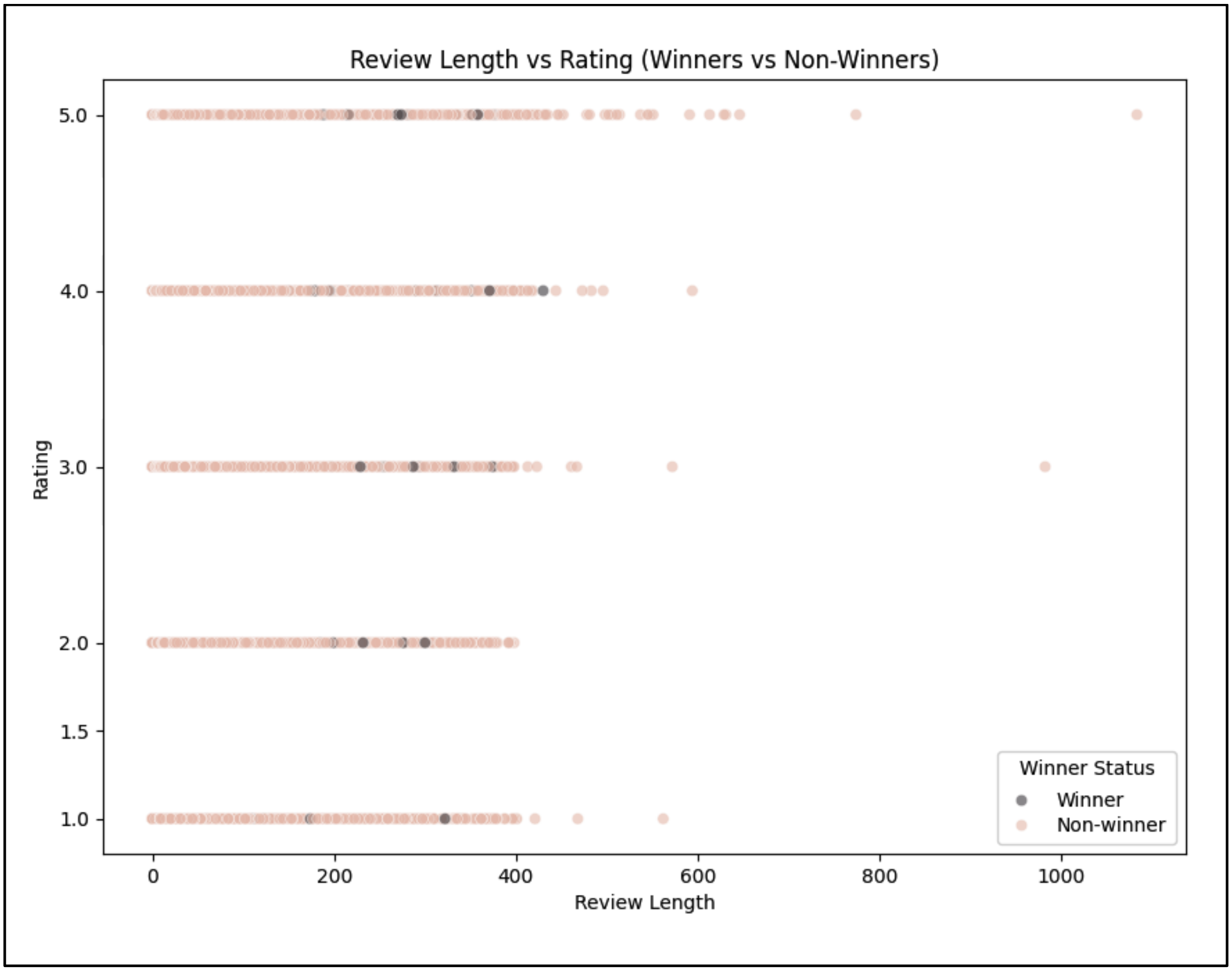

Figure 4 - A comparison of review length and ratings

3.2 Models

3.2.1 *Class Imbalance*

Class imbalance is a problem that arises in classification models when the classes involved are of different sizes. Classification models are generally built on the assumption that classes are evenly distributed (ScienceDirect, n.d.). However, in class-imbalanced datasets, these models often exhibit bias toward the majority class, resulting in inaccurate classification of the minority class and, potentially, poor overall performance (ScienceDirect, n.d.). In the case of the Sephora customer review data set, I web scraped over 778,000 records of customer review data from the Sephora website across over 7,000 products, 85 of which were Best of Beauty 2024 winners. This equates to just over 1% of the review records belonging to Best of Beauty winners.

My goal was to gather a large selection of review data for non-winners to provide the models with as much relevant data as possible for training. However, a significant challenge in the analysis was the limited pool of text available for training the models on winning products.

To address the class imbalance upfront, I reduced the non-winning review data, narrowing the TF-IDF data to about 800 non-winning products. Reducing the size of the non-winning data set had several benefits. First, the TF-IDF data was substantial. Every review resulted in 8,388,608 characters of TF-IDF metadata, stored as plain text in a Snowflake table. Running TF-IDF data on the reduced column set took considerable processing time. Using the full initial dataset would have been computationally unmanageable. A second benefit of reducing the dataset is that it minimized the class imbalance, and approximately 10% of the dataset was now composed of review data for winning products. A third benefit is that I narrowed the data set to exclude Allure Best of Beauty award winners from previous years. The initial data set included products that won awards in 2023 and earlier. To help improve model performance, I wanted a clear division in customer review data for winners and non-winners, so I removed the previous year's winners from the non-winner data set to ensure the non-winner data set encompassed only products that had never won a Best of Beauty award.

This analysis was a binary classification task to categorize customer ratings and review sentiment data into either a winner or non-winner category. The non-winning category represented the majority class, while the award-winning category was the minority class.

3.2.2 *TF-IDF*

To start, I used term frequency-inverse document frequency (TF-IDF) to quantify the importance of terms in Sephora customer review data. TF-IDF is a method for measuring the importance of a word in a document based on its frequency in that document (term frequency, TF) and its rarity across a collection of documents (inverse document frequency, IDF) (Capital One, n.d.). TF-IDF highlights important, distinctive terms while downplaying common words (Capital One, n.d.). TF-IDF is commonly used to analyze and rank textual data in document categorization, information retrieval, and keyword extraction. Machine learning algorithms work with numerical data, so text data must be converted into numbers through vectorization. TF-IDF vectorization calculates the importance of each word in a document and converts that information into a numerical vector. For this project, I used the TfidfVectorizer class from the scikit-learn library to transform customer review text into a matrix of TF-IDF features (Scikit-learn developers, 2024). By calculating these scores, I could numerically represent text data, which was then used to train machine learning models.

3.2.3 Sentiment Analysis (VADER)

To examine how customer sentiment differs between Best of Beauty award winners and non-award winners, I used Valence Aware Dictionary and sEntiment Reasoner (VADER). This pre-trained, rule-based sentiment analysis tool employs a pre-built lexicon to quantify sentiment in textual data. VADER is well-suited for customer reviews as it efficiently captures the emotional tone of text, providing quantitative sentiment scores through a compound score ranging from -1 (most negative) to +1 (most positive) (Slavanya, n.d.). VADER's simplicity and efficiency make it well-suited for analyzing large datasets, which is a top concern in this project. Its pre-trained, specific-word-based approach does not require model training, further saving time and computational resources while providing accurate sentiment scores.

3.2.4 Logistic Regression

I selected logistic regression as the first classification model to investigate how accurately we can predict whether a product will be a Best of Beauty award winner based on customer ratings and sentiment scores. I chose logistic regression for its simplicity, scalability, and computational efficiency, which were critical for a dataset of the size used in this project. The customer ratings table contained nearly 800,000 records of customer ratings. Two tables held the TF-IDF data: one table for winners and one table for non-winners. The largest of these tables is over 19.5 GB. As such, scalability and efficiency were top concerns when selecting a model.

Logistic regression assumes a linear relationship between the log odds of the dependent variable and the independent variables, which is reasonable given the numerical nature of the input features (e.g., sentiment scores and ratings). I created several logistic regression models to improve model accuracy and address the potential class imbalance, adjusting the class_weight parameter to "balanced" during the second iteration. I created the model using the train_test_split function from sklearn to split the data into training and testing sets and the LogisticRegressionclass from sklearn.linear_model to train the model. After making predictions on the test set, I used evaluation metrics such as accuracy, confusion matrix, and classification report generated using sklearn's classification_report, confusion_matrix, and accuracy_score functions. I performed hyperparameter tuning using 5-fold cross-validation.

3.2.5 Random Forest

I also employed Random Forest and XGBoost classification models to account for potential non-linear relationships in the data. Random Forest, an ensemble learning method, combines multiple decision trees to produce robust predictions (Scikit-learn Developers, n.d.). Using the RandomForestClassifier from sklearn.ensemble, I trained the model with 100 estimators (n_estimators) and set the class_weight parameter to "balanced" to address the class imbalance. I used grid search to fine-tune hyperparameters such as the maximum depth of trees (max_depth) and the minimum number of samples per split (min_samples_split). I also applied a 5-fold cross-validation approach to validate the model.

3.2.5 XGBoost

The last model I trained was XGBoost, a gradient-boosting algorithm selected for its speed, scalability, and effectiveness in handling large datasets and class imbalances (XGBoost Developers, n.d.). I implemented the model using the xgboost library in Python, splitting the data into training and testing sets and training the model on the training set. I used grid search to train the hyperparameters such as learning rate (eta), maximum tree depth (max_depth), and the number of boosting rounds. Using XGBoost, I adjusted the parameter scale_pos_weight to handle class imbalance (XGBoost Developers, n.d.). Again, I used sklearn's metrics module for model evaluation, including calculating accuracy, precision, recall, F1 score, and AUC-ROC, and used cross-validation to evaluate model performance.

4 ANALYSIS AND RESULTS

This study aimed to answer three research questions analyzing customer sentiment in relation to the 2024 Best of Beauty award winners by examining Sephora customer reviews for winners and non-winners.

Are there specific keywords in customer reviews that distinguish Best of Beauty winners from non-winners?

Using TF-IDF to identify the top 50 words for both winners and non-winners showed little crossover in the top ranked words between the two categories. Of the top 50 words in each category, only 13 words could be found on both lists. The 13 words that exist in both lists are: bit, dark, found, kinda, looking, lot, perfectly, practical, setting, soft, strikes, sunscreen, and two.

Figure 5 - Top 50 words in customer reviews for Best of Beauty winners (as determined by TF-IDF)

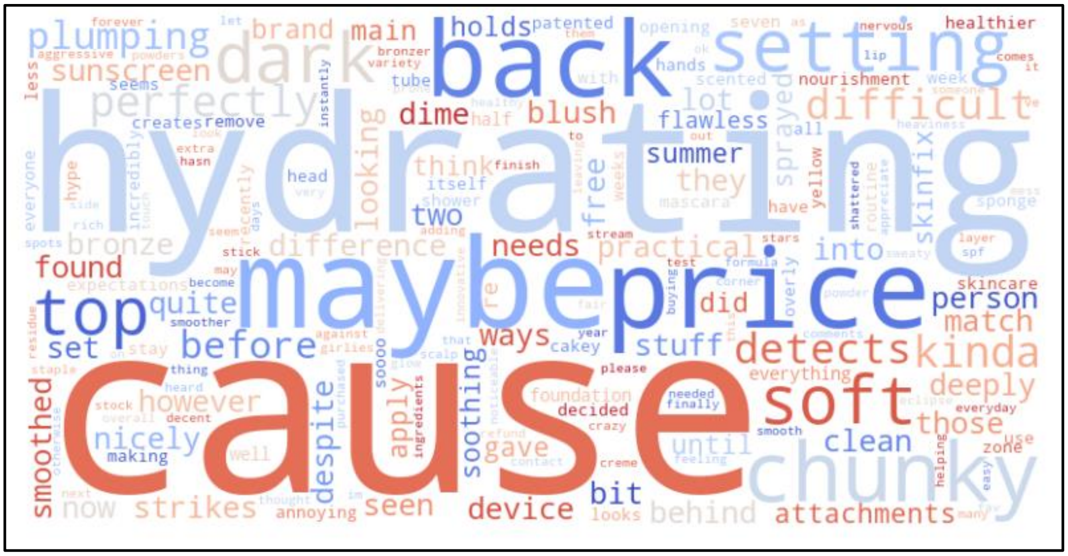

Figure 6 - Top 50 words in customer reviews for non-winners (as determined by TF-IDF)

How does customer sentiment differ between Best of Beauty award winners and non-award winners?

Using VADER to quantify text sentiment, we find that the sentiment score is statistically significantly higher for Best of Beauty award winners. A t-statistic of 4.2714 indicates a significant difference between the two groups, while a P-value of 0.000 confirms that this difference is statistically significant. The density plot, shown in Figure 7, shows that the sentiment is generally higher for winners, which is consistent with the statistical results.

The VADER analysis indicated a significant difference in customer sentiment between winning and non-winning products. However, the statistical significance does not imply that sentiment alone is a sufficient predictor for determining which products are winners

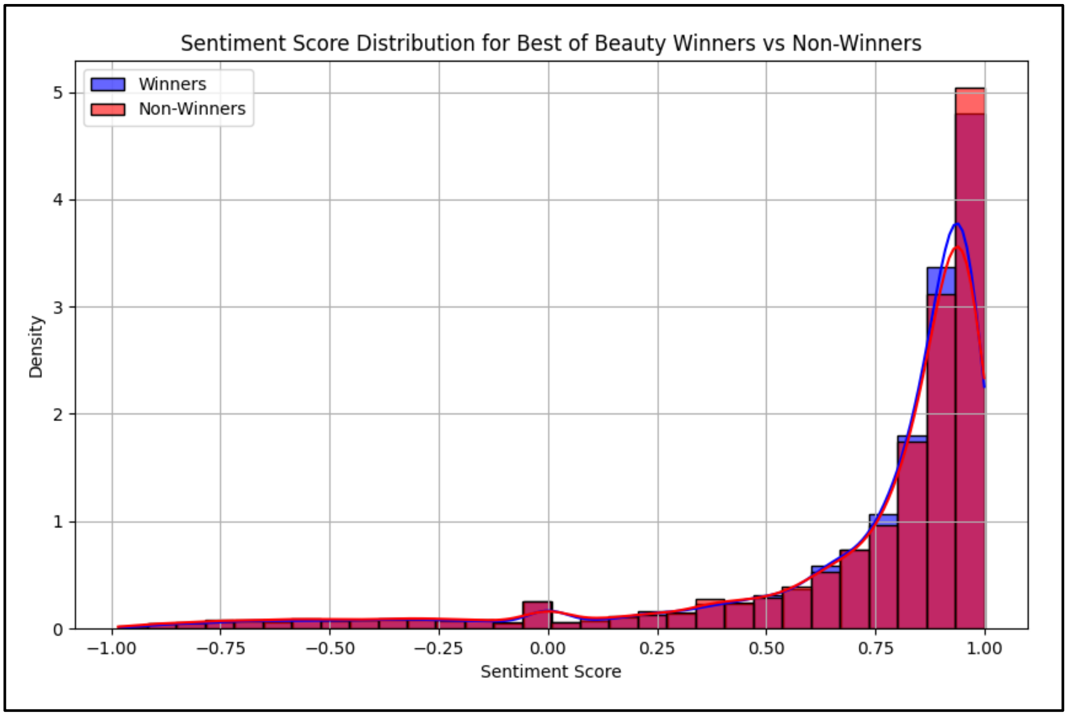

Figure 7 - Sentiment Analysis Results (VADER)

How accurately can we predict which products will be Best of Beauty winners based on customer ratings and sentiment scores? Can we use specific keywords in customer reviews that distinguish Best of Beauty winners from non-winners help predict future award winners?

Despite training and testing a range of classification models, including logistic regression, Random Forest, and XGBoost, tuning parameters, and taking steps to accommodate class imbalance, the features I trained the models on—sentiment scores, ratings, and keyword analysis - the core components of a customer review - were not effective at accurately predicting Best of Beauty winners. The models consistently struggled with classifying correctly, highlighting a strong bias toward predicting the majority class and achieving only moderate overall performance.

The models and their respective prediction accuracy are shown below in Table 4.

Model	Accuracy
Logistic Regression (prior to adjusting for class imbalance)	58.5%
Logistic Regression (after adjusting for class imbalance)	56%
Random Forest	52.5%
XGBoost	52.9%

Table 4 **-** Classification model performance

4.1 Classification Model Analysis

After training the first logistic regression model, the test data set showed an accuracy of 58.5%. The model performance also showed significant class imbalance issues, accurately identifying the majority class at 98%. After adjusting for class imbalance, the second Logistic Regression model showed a prediction accuracy of 56%. It improved on correctly predicting the minority class (up from 3% to 38%) but at the expense of accurately predicting the majority class, which dropped to 69% accuracy.

After training and testing, the Random Forest model performed with an overall accuracy of 52.5%. The results were more balanced than the Logistic Regression model, accurately predicting the minority class at 47%. This confirms that the Random Forest model handles the class imbalance better, but still performed with an overall low accuracy.

The XGBoost model predicted winners and non-winners with 52.9% accuracy, showing the best recall for the minority class of all the classification models (48%). Overall performance was comparable to the other classification models.

4.2 Overlap Between the Winners and Non-Winners

Even though the VADER analysis showed that winners tend to have higher sentiment scores and customer ratings on average, there is still a significant overlap in sentiment scores between winners and non-winners, as both are generally high, which makes it difficult for models to reliably separate the two groups.

5 CONCLUSION

This analysis demonstrates that customer review feedback, comprised of ratings (1-5), review text, and sentiment analysis, is insufficient for predicting Allure Best of Beauty award winners. Generally, customer reviews skew towards positive sentiment and higher ratings. Still, there is a statistically significant increase in positive customer sentiment towards award-winning products versus non-winners.

5.1 Actionable Results

This analysis highlights that while customer reviews—comprising ratings, text, and sentiment—do not predict Allure Best of Beauty winners, award-winning products show statistically significantly higher positive sentiment. Brands can use this by enhancing product quality, addressing recurring themes in customer reviews, and incorporating positive sentiment drivers into marketing campaigns.

Although specific keywords don't guarantee awards, as the classification models show, aligning product messaging with top-reviewed attributes can resonate with customers. Below are the top 50 keywords identified by TF-IDF found in the customer reviews for Best of Beauty award winners.

1.	maybe	14.	side	27.	rather	40.	helped
2.	yellow	15.	along	28.	fav	41.	dark
3.	strikes	16.	experienced	29.	mask	42.	slight
4.	purchased	17.	refund	30.	gave	43.	wore
5.	main	18.	recommend	31.	gets	44.	bit
6.	how	19.	kinda	32.	practical	45.	happy
7.	compare	20.	dyson	33.	game	46.	looking
8.	found	21.	lot	34.	giving	47.	try
9.	as	22.	flake	35.	skincare	48.	washing
10.	sunscreen	23.	setting	36.	only	49.	Instead
11.	perfectly	24.	of	37.	innovative	50.	causing
12.	two	25.	lip	38.	dispensing		
13.	soft	26.	helping	39.	tzone		

Figure 8 - Top 50 words in customer review text for Best of Beauty winners

After winning an award, brands should also emphasize the award recognition in marketing to boost credibility and capitalize on the potential for increased sales.

5.2 Lessons Learned

In this analysis, the size of the dataset exceeded initial expectations, which created challenges in terms of computational resources required to train and run the models effectively. To address this, I created a Snowflake trial account to utilize cloud-based storage and processing resources, eliminating the need to store the dataset locally. I also switched from R to Python to improve processing efficiency, as several models required hours to complete successfully.

I incorporated error handling and intermittent data storage to mitigate issues with long runtimes and potential interruptions. For example, I stored data in Snowflake temporary tables every 50 records when calculating TF-IDF values. This approach ensured that I retained already processed values even if the script encountered errors or stopped running, allowing the analysis to resume without restarting the entire process. Additionally, I broke down complex tasks into smaller, more manageable steps. For instance, when collecting customer reviews, I limited the script to handle 100-250 product IDs at a time.

This project emphasized the importance of adapting to accommodate the demands of large datasets, particularly when working with limited computational resources.

6 REFERENCES

1. Capital One. (n.d.). *Understanding TF-IDF*. Capital One Tech. https://www.capitalone.com/tech/machine-learning/understanding-tf-idf/

2. Chia, Jessica. (2019, August 15). *The truth about online beauty product reviews: Can you trust them?*. Allure. https://www.allure.com/story/trust-online-beauty-product-reviews-truth

3. Cosmetify. (n.d.). *The Cosmetify Index*. https://www.cosmetify.com/us/the-cosmetify-index/

4. FasterCapital. (2024, June 15). *Beauty award nomination: Behind the scenes—How beauty brands secure award nominations*. https://fastercapital.com/content/Beauty-award-nomination--Behind-the-Scenes--How-Beauty-Brands-Secure-Award-Nominations.html

5. Federal Trade Commission. (2019a, October). *Devumi owner and CEO settle FTC charges they sold fake indicators of social media influence to cosmetics firm.* https://www.ftc.gov/news-events/news/press-releases/2019/10/devumi-owner-ceo-settle-ftc-charges-they-sold-fake-indicators-social-media-influence-cosmetics-firm

6. Federal Trade Commission. (2019b, October 21). *Complaint: In the matter of Sunday Riley Modern Skincare, LLC.*https://www.ftc.gov/system/files/documents/cases/192_3008_c4729_sunday_riley_complaint.pdf

7. Federal Trade Commission. (2020, November). *FTC approves final consent agreement with Sunday Riley Modern Skincare, LLC.* https://www.ftc.gov/news-events/news/press-releases/2020/11/ftc-approves-final-consent-agreement-sunday-riley-modern-skincare-llc

8. Federal Trade Commission. (2024, August 14). *Federal Trade Commission announces final rule banning fake reviews and testimonials.* Updated September 24, 2024. https://www.ftc.gov/news-events/news/press-releases/2024/08/federal-trade-commission-announces-final-rule-banning-fake-reviews-testimonials

9. Inside Sephora. (n.d.). *Life at Sephora*. https://www.inside-sephora.com/en/usa/life-at-sephora

10. LinkedIn. (n.d.). *Which machine learning algorithms provide the fastest analysis?* https://www.linkedin.com/advice/3/which-machine-learning-algorithms-provide-fastest-nsq5f#:~:text=Linear%20algorithms%20such%20as%20Linear,significantly%20speeds%20up%20the%20analysis.

11. PowerReviews. (2023a). *Health & beauty shopping trends 2023*. https://www.powerreviews.com/research/health-beauty-shopping-trends-2023/

12. PowerReviews. (2023b). *Survey: The ever-growing power of reviews (2023 edition).* https://www.powerreviews.com/power-of-reviews-2023/

13. ScienceDirect. (n.d.). *Classification models.* https://www.sciencedirect.com/topics/computer-science/classification-models

14. Scikit-learn developers. (2024a). *sklearn.feature_extraction.text.TfidfVectorizer — Scikit-learn 1.5.0 documentation*. https://scikit-learn.org/1.5/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

15. Scikit-learn Developers. (n.d.). *sklearn.ensemble.RandomForestClassifier.* https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

16. Slavanya, R. (n.d.). *VADER: A comprehensive guide to sentiment analysis in Python*. Medium. https://medium.com/@rslavanyageetha/vader-a-comprehensive-guide-to-sentiment-analysis-in-python-c4f1868b0d2e

17. XGBoost Developers. (n.d.). *XGBoost documentation.* https://xgboost.readthedocs.io/en/stable/