**Sentiment Analysis of Amazon Beauty Product Reviews**

(Group 1 Members)

Fatimah Aljohani, Aishwarya Shastry Viswanath and Mi Kin Swan

Drexel University, Philadelphia, PA

DSCI 521: Data Analysis and Interpretation

Instructor: Dr. Pragati Awasthi

04/20/2025

# Abstract

# This project explores sentiment analysis of Amazon customer reviews in the "All Beauty" product category. The primary goal is to understand customer opinions and how they correlate with product features such as price, category, and verified purchase status. We utilized both review text and metadata to perform exploratory data analysis (EDA) and classify sentiment.

# We began by cleaning and merging two datasets: one containing raw reviews and another with product metadata. Due to incomplete category data, we implemented a custom sub-category classification based on product titles and descriptions to label products (e.g., Hair Care, Skin Care, Makeup).

# The analysis revealed that most reviews are positive, verified purchases are slightly more favorable, and sentiment trends vary across sub-categories and pricing ranges. This foundational work sets the stage for building a sentiment classification model in future phases, with potential applications for e-commerce insight, product improvement, and customer feedback monitoring.

# Team Introduction

Our team consists of three students with diverse expertise in Python programming, data acquisition, data wrangling, and visualization.

1. **Fatimah Aljohani**: In this project, I was responsible for collecting and preparing the dataset. I combined Amazon product reviews with corresponding metadata, handled the cleaning and preprocessing of the data, and engineered new features. I have prior experience with sentiment analysis in my native language, which helped me understand the challenges of working with text data. This background allowed me to prepare the data for future modeling and evaluation.
2. **Aishwarya Shastry Viswanath:** In this project, I was responsible for data cleaning, preprocessing, and feature engineering. I added new features such as product sub-categories, review length, sentiment polarity scores, and helpfulness ratios to enhance model input and analysis depth. I performed Exploratory Data Analysis using visualizations like rating distributions, word clouds, and sub-category-wise sentiment trends to identify patterns in customer feedback. I applied Natural Language Processing (NLP) techniques to preprocess text data and build a sentiment classification model to categorize reviews as positive, neutral, or negative. This project strengthened my skills in text analytics, feature creation, and data visualization, while also deepening my understanding of consumer behavior in the beauty product domain.
3. **Mi Kin Swan:** In this project, my role involved cleaning the raw review texts by removing special characters, converting text to lowercase, eliminating stopwords, and performing tokenization to prepare the data for sentiment analysis. I also conducted exploratory data analysis (EDA) to gain insights into the dataset. Moving forward, I am interested in improving my skills in Natural Language Processing (NLP), particularly in using advanced techniques such as word embeddings and deep learning models. I also aim to become more confident in every step of the data science life cycle to further strengthen my understanding and contribute more effectively to future projects.

# **Dataset Description**

The dataset utilized for this project is the Amazon Reviews 2023 dataset, compiled by the McAuley Lab and hosted on Hugging Face. This extensive dataset includes over 571 million customer reviews from May 1996 to September 2023, spanning 33 product categories. The dataset is divided into two parts: UserReviews and ItemMetadata. For the purposes of this project, we focus specifically on the "All Beauty" category. To enrich the analysis, we joined the two parts to obtain additional product-related information. After merging, the resulting dataset contains around 701,528 rows and 27 columns.

### **For User Reviews Attributes and Datatype**

|  |  |  |
| --- | --- | --- |
| **Field** | **Type** | **Explanation** |
| rating | float | Rating of the product (from 1.0 to 5.0). |
| title | str | Title of the user review. |
| text | str | Text body of the user review. |
| images | list | Images that users post after they have received the product. Each image has different sizes (small, medium, large), represented by the small\_image\_url, medium\_image\_url, and large\_image\_url respectively. |
| asin | str | ID of the product. |
| parent\_asin | str | Parent ID of the product. Note: Products with different colors, styles, sizes usually belong to the same parent ID. The “asin” in previous Amazon datasets is actually parent ID. Please use parent ID to find product meta. |
| user\_id | str | ID of the reviewer |
| timestamp | int | Time of the review (unix time) |
| verified\_purchase | bool | User purchase verification |
| helpful\_vote | int | Helpful votes of the review |

### **For Item Metadata Attributes and DataType**

|  |  |  |
| --- | --- | --- |
| **Field** | **Type** | **Explanation** |
| main\_category | str | Main category (i.e., domain) of the product. |
| title | str | Name of the product. |
| average\_rating | float | Rating of the product shown on the product page. |
| rating\_number | int | Number of ratings in the product. |
| features | list | Bullet-point format features of the product. |
| description | list | Description of the product. |
| price | float | Price in US dollars (at time of crawling). |
| images | list | Images of the product. Each image has different sizes (thumb, large, hi\_res). The “variant” field shows the position of image. |
| videos | list | Videos of the product including title and url. |
| store | str | Store name of the product. |
| categories | list | Hierarchical categories of the product. |
| details | dict | Product details, including materials, brand, sizes, etc. |
| parent\_asin | str | Parent ID of the product. |
| bought\_together | list | Recommended bundles from the websites. |

This dataset is particularly suitable for sentiment analysis as it provides both unstructured text (review content) and structured feedback (numerical ratings), allowing for effective labeling of sentiment (e.g., positive, neutral, negative). The large volume and diversity of reviews support robust machine learning and natural language processing techniques.

# **Exploratory Questions**

To guide our analysis, we developed several exploratory questions to better understand customer sentiment and product characteristics in the All-Beauty category. The following questions were directly addressed through code and visualizations in our notebook:

1. What is the overall distribution of star ratings?
2. How is sentiment (positive, neutral, negative) distributed across the dataset?
3. How does sentiment differ between verified and unverified purchases?
4. What is the relationship between product price and review sentiment?
5. What are the most frequent words used in positive vs. negative reviews (via WordCloud)?
6. How many reviews are posted over time (monthly trend)?
7. Which product sub-categories (e.g., Hair Care, Skin Care, Makeup) receive the most reviews?

# **EDA Summary**

# After collecting the data, we merged two datasets — UserReview and ItemMetadata — and renamed column names to make them more descriptive and easier to work with. We then performed preprocessing steps such as cleaning text data and handling missing values.

In the following visualizations, we highlight the key aspects we explored during our EDA of Amazon beauty product reviews. These charts helped us uncover trends in customer sentiment, product types, and purchasing behavior.

### Distribution of Star Ratings

We checked the distribution of review ratings. To prepare for sentiment analysis, we labeled the ratings as follows:

* Positive: ratings ≥ 4
* Neutral: ratings: rating =3
* Negative: <= 2

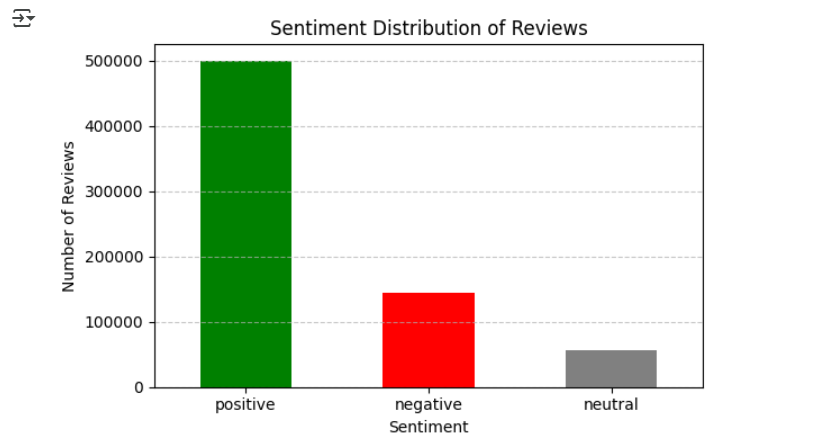


Figure 1. Distribution of Sentiment Labels Based on Star Ratings

The bar chart shows that the majority of reviews were labeled as positive (rating ≥ 4), making up more than half of the dataset. Negative reviews (rating ≤ 2) represent a smaller but still significant portion, while neutral reviews (rating = 3) were the least common. This distribution highlights a positive skew in customer feedback, which is typical for beauty product reviews where satisfied customers are more likely to leave comments.

### Original Distribution of Star Ratings

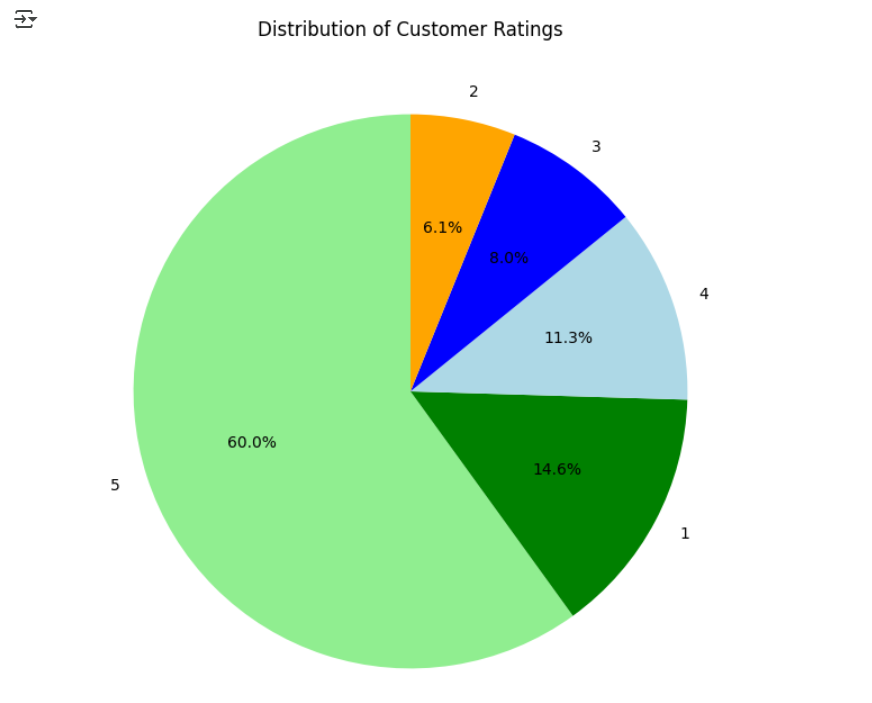


Figure 2. Pie Chart Showing the Original Customer Rating Breakdown (1 to 5 stars)

This pie chart displays the raw distribution of star ratings from 1 to 5. The majority of reviews are rated 5 stars, accounting for 60% of the dataset. Interestingly, the second most common rating is 1 star (14.6%), which suggests a polarized customer experience—many users are either highly satisfied or very disappointed. Ratings of 4, 3, and 2 stars make up smaller portions, indicating that moderate reviews are less frequent.

### Sentiment Distribution by Verified Purchase Status

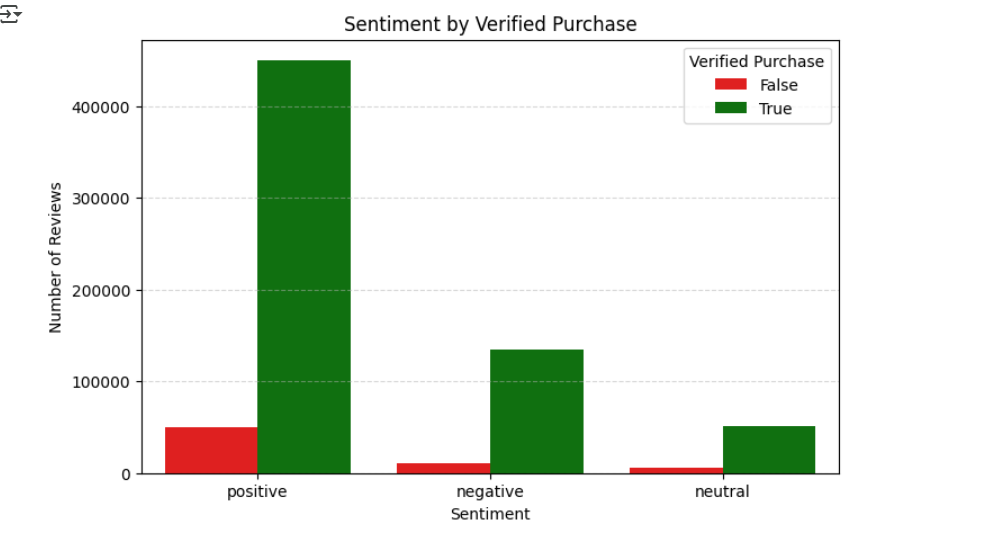
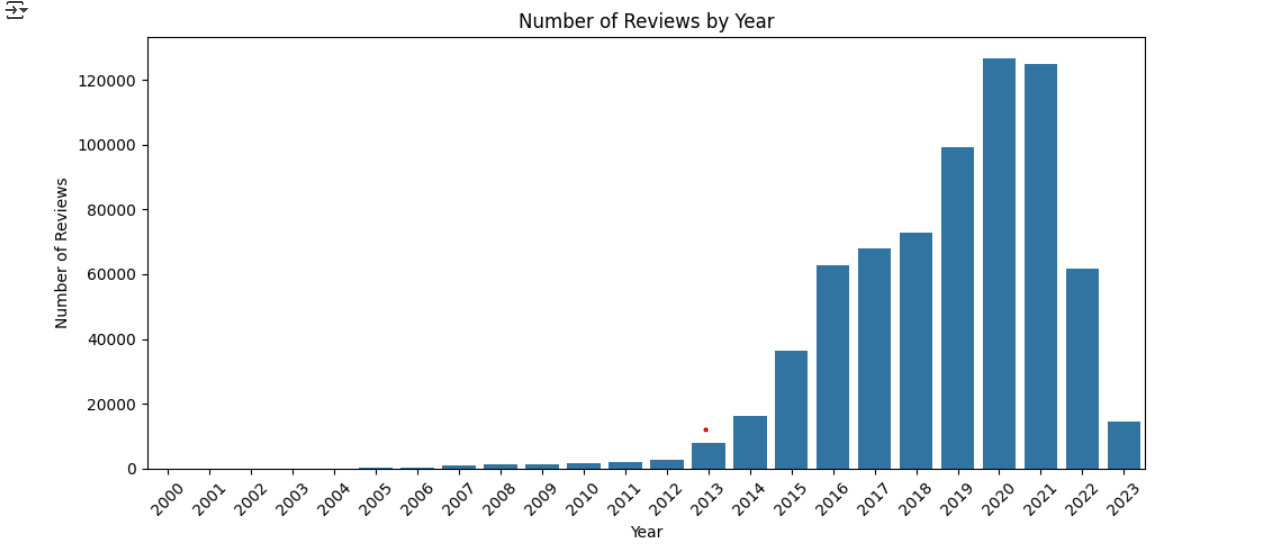


Figure 3. Sentiment Distribution Among Verified and Unverified Purchases

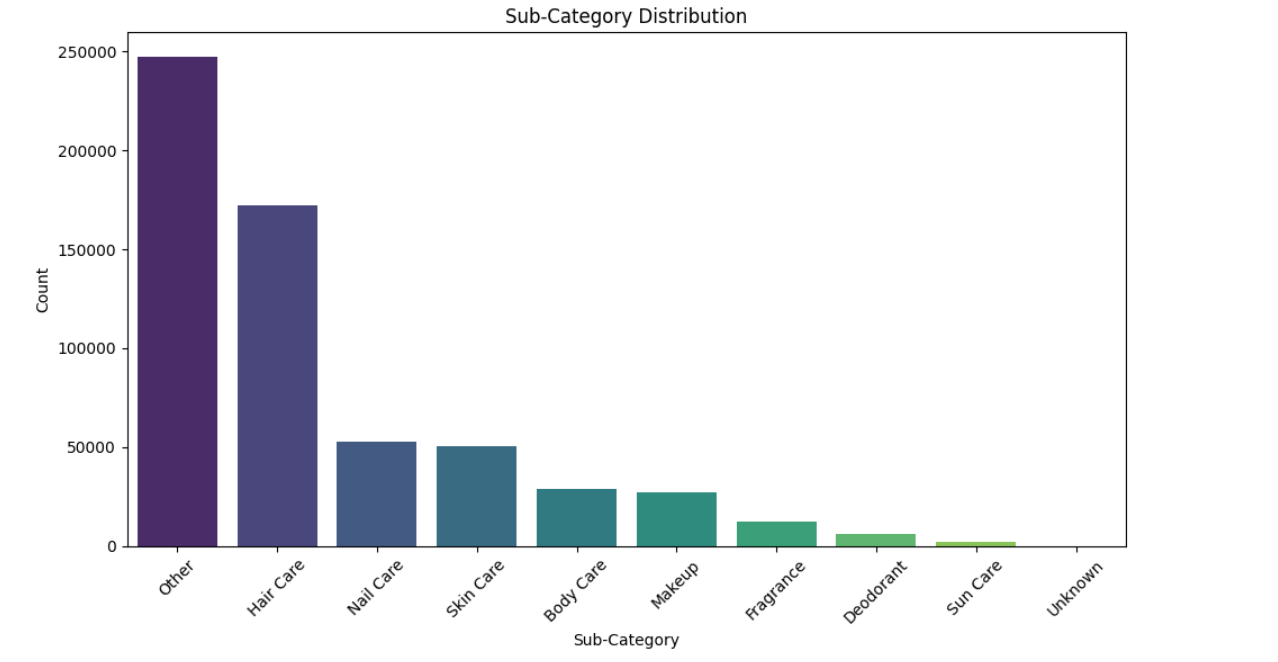
This bar chart illustrates the distribution of sentiment labels across verified and unverified purchases. Most reviews, **especially positive ones,** come from verified purchases, indicating a strong correlation between actual product experience and favorable feedback. In contrast, unverified purchases show a relatively higher proportion of negative sentiment, which may suggest that customers who haven't genuinely experienced the product are more likely to leave critical or misleading reviews.

### Annual Trend of Customer Reviews

Figure 4. Number of Customer Reviews by Year

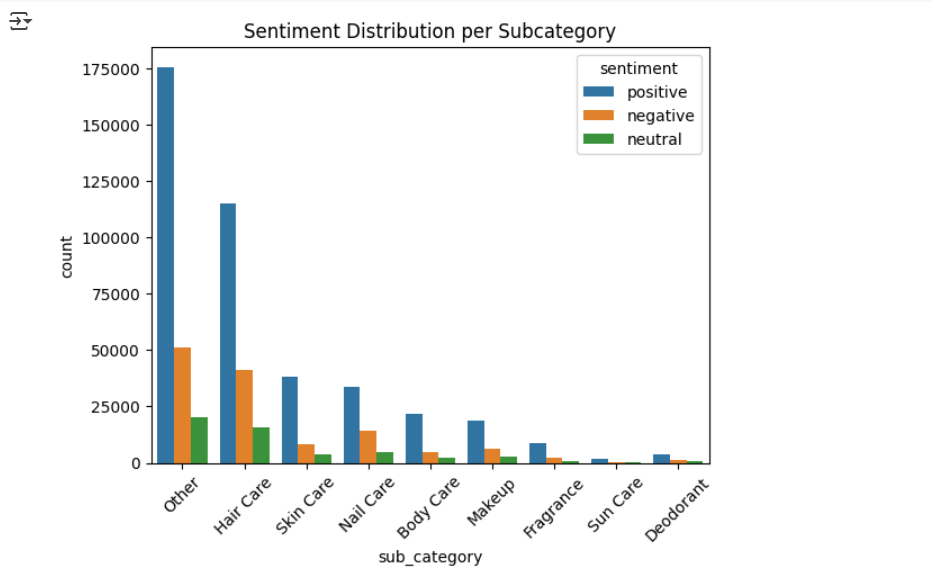
The bar chart shows the number of product reviews submitted each year. Review activity began to increase noticeably around 2013 and continued to grow steadily until it peaked in 2020 and 2021. This surge could be attributed to increased online shopping during the COVID-19 pandemic. After 2021, the number of reviews declined, possibly due to reduced consumer activity or incomplete data for the most recent years.

### Custom Sub-Category Classification of Beauty Products

Figure 5. Distribution of Reviews Across Custom Sub-Categories

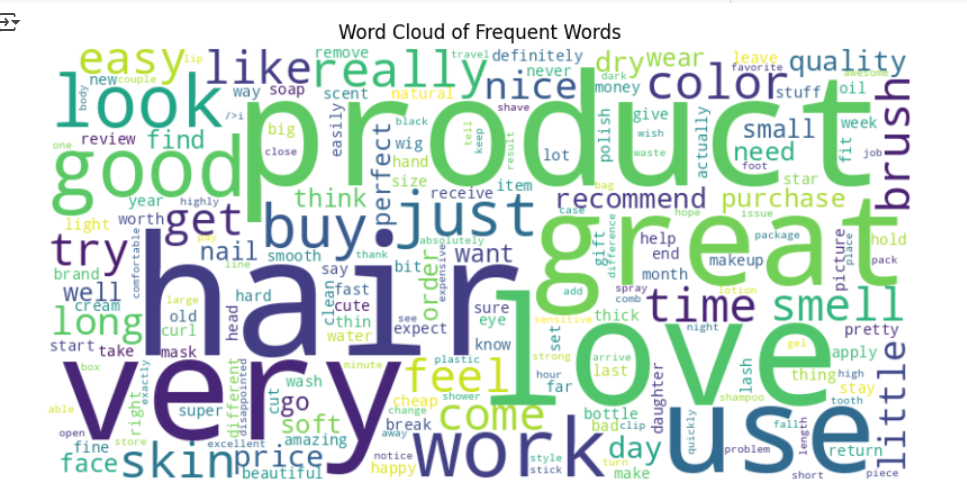
This bar chart presents the distribution of beauty products across custom sub-categories that we generated using keyword extraction from product titles and descriptions. “Hair Care” had the highest number of clearly categorized products, followed by “Nail Care” and “Skin Care.” A significant portion of the products could not be confidently classified, resulting in the “Other” category. This highlights the challenges of working with unstructured product data and justifies the importance of building a rule-based classification approach.

### Sentiment Distribution Across Product Sub-Categories

Figure 6. Sentiment Breakdown per Sub-Category

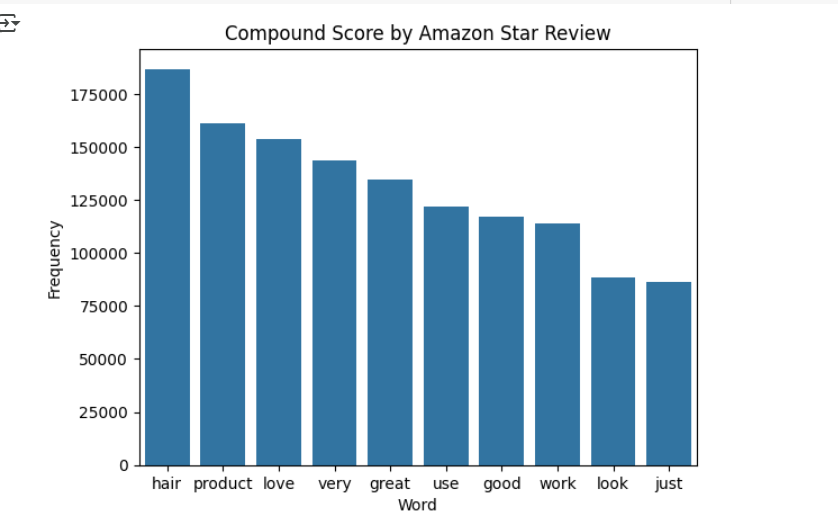
This bar chart illustrates how customer sentiment is distributed within each of the custom sub-categories. Most sub-categories are dominated by positive reviews, especially in Hair Care and Skin Care. However, some categories like Nail Care and Fragrance show relatively higher proportions of negative feedback. These differences may reflect variations in product effectiveness, customer expectations, or sensitivity of the product type. This insight is valuable for brands looking to identify which product lines need quality improvement or reputation management.

### Commonly Used Words in Product Reviews

Figure 7. Word Cloud of Most Frequent Words in Customer Reviews

The word cloud highlights the most frequently used words in the product review texts. Words such as “product,” “love,” “hair,” “great,” and “use” dominate the visual, suggesting that customers often express affection and satisfaction in their reviews. Other words like “very,” “good,” “buy,” and “work” further reinforce the positive sentiment seen in the earlier analysis. This visualization provides a quick overview of customer language and confirms the generally favorable tone present throughout the dataset.

### Most Frequently Used Words in Reviews

Figure 8. Top 10 Most Frequent Words in Review Texts

This bar chart shows the top 10 most commonly used words in the review texts. The most frequent word is “hair,” followed by “product,” “love,” “very,” and “great.” The presence of these terms highlights the focus on hair-related products in the “All Beauty” category. The dominance of positive words like “love,” “great,” and “good” also aligns with the strong positive sentiment observed throughout the dataset. These word frequencies offer a clear summary of common topics and emotional tones in customer feedback.

# **Data selection and Justification**

We selected the Amazon Reviews 2023 dataset from the McAuley Lab (via Hugging Face) and focused on the “All-Beauty” category. This dataset includes over 700,000 product reviews along with structured metadata such as product titles, prices, descriptions, verification status, timestamps, and star ratings.

One of our main goals in this project was to improve our skills in Natural Language Processing (NLP), particularly in text preprocessing, sentiment classification, and word pattern analysis. The Amazon Beauty review data is well-suited for this purpose due to the richness and expressiveness of the review text.

Additional reasons for choosing this dataset include:

* Combination of review text and product metadata, which enables deeper analysis by connecting sentiment with product features like price and verification status.
* Clear and high-quality data, suitable for both EDA and supervised learning tasks.
* Scalability, with a large sample size that allows for meaningful statistical insights.
* Custom categorization potential, which allowed us to create sub-categories (e.g., Hair Care, Skin Care, Makeup) by analyzing product titles and descriptions due to the lack of detailed category labels in the original data.

Overall, this dataset provided a strong foundation for learning and applying NLP techniques in a realistic and valuable context.

# **Analysis Completion and Dissemination plan**

The analysis will begin with preprocessing the dataset, which includes cleaning the text data, handling missing values, removing duplicates, and tokenizing the review texts. After preprocessing, the reviews will be labeled as positive, neutral, or negative based on their star ratings to create a target variable for sentiment classification. Natural Language Processing (NLP) techniques such as TF-IDF vectorization or word embeddings will be applied to convert the textual data into numerical form. Then, various machine learning models will be trained and evaluated to determine the best-performing classifier for predicting sentiment.

For dissemination, the findings will be presented through a combination of:

* Visualizations (e.g., word clouds, sentiment distribution charts)
* A documentation file and README to explain the methodology, results, and how to reproduce the analysis

This approach will ensure that both technical and non-technical audiences can interpret the results and gain insights into the sentiment trends in the “All Beauty” product reviews.

# **Application / Relevance of Analysis**

The sentiment analysis of Amazon Beauty product reviews holds practical value in transforming unstructured customer feedback into actionable insights. By categorizing reviews into positive and negative sentiments, businesses can quickly assess product performance, customer satisfaction, and emerging trends. This analysis helps brands identify areas for improvement, understand consumer preferences, and refine product offerings accordingly. It also aids marketing teams in designing targeted campaigns based on sentiment trends and e-commerce platforms in improving product ranking algorithms and personalized recommendations. Overall, the model bridges the gap between customer voice and business strategy, enabling data-driven decision-making across the beauty product lifecycle.

# **Target Audience**

The target audience for this sentiment analysis includes stakeholders in the beauty and e-commerce industry who are invested in understanding customer opinions and enhancing product satisfaction. Specifically:

* Beauty Product Manufacturers To understand real customer opinions and improve product design, formulation, and quality based on user experiences.
* Marketing and Brand Managers To monitor customer sentiment trends and develop targeted advertising campaigns that resonate with consumer needs and preferences
* E-commerce Platforms: To enhance product recommendation systems, improve review filtering, and better serve shoppers by surfacing relevant feedback.
* Customer Service Teams: To proactively address common issues highlighted in negative reviews and improve overall customer satisfaction.
* Business Analysts and Product Managers:To track performance across different beauty products and categories, aiding in strategic planning, inventory management, and pricing.

# **Limitations of Analysis**

While sentiment analysis offers valuable insights into consumer opinions, there are several notable limitations in analyzing Amazon beauty product reviews:

* Implicit Sentiments: Many reviews express opinions subtly. Phrases like "Item as described" or "It works" may appear neutral but imply satisfaction or disappointment depending on context, which can be misclassified by sentiment models.
* Sarcasm and Irony: Models often fail to detect sarcasm or irony, especially without tone or context. A review like "Exactly what I needed—another useless product" may be misinterpreted as positive or neutral when it's actually negative.
* Domain-Specific Jargon and Cultural References: Beauty products often involve terminology and cultural nuances (e.g., undertones, texture, blendability) that generic sentiment models may not accurately interpret. This can lead to misclassification of sentiment.
* Lack of Fine-Grained Sentiment Analysis: Many models operate on a binary or ternary scale (positive, neutral, negative), failing to capture more nuanced opinions such as sentiment toward specific product attributes (e.g., packaging vs. effectiveness).

# Conclusion

This project provided valuable insights into customer sentiment on Amazon beauty products by analyzing reviews alongside product metadata. We discovered that most reviews are positive, especially for verified purchases, and that sentiment patterns vary across product types such as Hair Care and Fragrance. Custom sub-category classification also helped overcome limitations in the original data.

In future work, we aim to build a sentiment classification model using review text. This will involve preparing the text data, applying machine learning techniques, and evaluating the model’s performance to support more scalable and automated insights.