The Evolving Beauty Industry

Balancing Safety, Pricing, and Consumer Trust

By: Kristine Zhou, Cedric Lambert, Yingying Sun

In today's world, the skincare and beauty products we use every day are under increasing scrutiny. From the moisturizer applied each morning to the lipstick chosen for a night out, the ingredients within these products have sparked intense discussion. Consumers are no longer passive users; they demand safer, cleaner formulations, prompting brands to scramble in response to ever-evolving regulatory standards. Behind the glossy packaging lies a complex web of challenges: how do brands maintain competitive pricing while ensuring their products meet safety standards? What strategies do they employ to preserve positive average ratings in a market increasingly focused on ingredient transparency?

The beauty industry has long walked a tightrope between science and perception. A single study or regulation can transform a prized ingredient into something to be avoided. As consumer awareness rises, the trend toward "clean beauty" intensifies, placing immense pressure on brands to reformulate their products and discontinue ingredients that no longer meet safety expectations. This leads us to critical questions: How quickly do brands adapt to these shifts, and how do they balance innovation, safety, and consumer trust?

Introduction:

Our choice to explore this area is rooted in the broader context of regulatory influence on market behavior. Recent changes in regulations, such as the banning of phthalates and parabens in cosmetics, have reshaped product formulations. Reports from organizations like the Environmental Working Group (EWG) and changes in California's Safe Cosmetics Program illustrate how regulatory bodies are actively pushing for safer formulations. According to Lignopure.com: "*Cosmetic manufacturers are required to provide an ingredient list on the product packaging or label which typically lists the ingredients in descending order of concentration, with the most abundant ingredients listed first. This allows consumers to understand what the product contains and make informed decisions based on their preferences or sensitivities.*"(Lignopure, 2024)[1]. This level of transparency is crucial in building consumer trust, as individuals increasingly seek brands that align with their values and safety expectations. Brands that prioritize clear labeling and ingredient disclosure are more likely to foster loyalty and stand out in a competitive market. Surveys indicate that 60% of consumers expect brands to disclose the source of their ingredients, while 72% want clear explanations of what ingredients do.(Cosmetics Business, 2024)[2]. At the same time, according to cleanhub.com: "*63% of consumers deem clean beauty as 'extremely or very important' when selecting cosmetics, showcasing a robust inclination towards environmentally conscious brands.*" (Cleanhub, 2023)[3]. These findings underscore the importance of transparency in

building trust and loyalty among today's consumers. Our project aims to delve deep into a comprehensive dataset that encapsulates a wealth of information about beauty brands and their products, uncovering insights into brand strategies and consumer preferences.

At the heart of our analysis lies a suite of key metrics that illuminate brand behaviors. We will explore product counts to quantify the breadth of each brand's offerings, providing insight into their market presence. Discontinuation rates will be scrutinized to understand which brands streamline their lines and the factors driving these decisions. Ingredient removals will also be a focal point, highlighting how brands respond to regulatory pressures and changing consumer sentiments regarding product safety.

Additionally, we will analyze average concentrations of ingredients, offering insights into formulation trends. By assessing how brands incorporate or eliminate certain ingredients, we can discern broader patterns in product development and consumer preferences.

Finally, we will examine market performance indicators such as average prices and customer ratings. In a competitive landscape where brands set high standards for inclusivity and ingredient safety, understanding these dynamics is crucial. By correlating these metrics, we aim to unravel the relationships between pricing strategies, product quality, and consumer satisfaction, ultimately shedding light on how these various factors interplay to shape brand identities and consumer experiences.

Our findings could inform future product development strategies, guiding brands in crafting formulations that meet current regulatory standards while anticipating future consumer expectations.

Data Sources:

The CDPH Product and Ingredient Dataset, located at https://cscpsearch.cdph.ca.gov/search/publicsearch, is a comprehensive collection that provides detailed information on cosmetic products sold in California, focusing on their ingredients and safety assessments. This dataset is provided in csv format and contains essential variables such as Brand, Product Category, Product Discontinued Date, Product Submitted Date, Ingredient Removed Date, and Concentration. It comprises 718660 observations (total 229 MB) and covers submissions and removals from June 2009 to September 2024. This dataset serves as a critical resource for understanding how regulatory changes impact product formulations and consumer safety in the cosmetics industry.

The Top Beauty & Cosmetics Products Worldwide 2024 dataset is accessible at https://www.kaggle.com/datasets/waqi786/most-used-beauty-cosmetics-products-in-the-world/data and offers insights into global beauty and cosmetic product trends, including various attributes of the products. This dataset is formatted as csv and includes key variables such as

Brand, Price_USD, Rating, and Number_of_Reviews. It contains 15000 observations (total 1.57 MB) and specifically focuses on products for the year 2024. This dataset is valuable for analyzing market trends and consumer preferences in the beauty sector, allowing for a comparison of pricing strategies and product ratings across different brands.

Data Manipulation Methods

Analysis

Analyzing Impact of Products Reported(09 Thru 24)

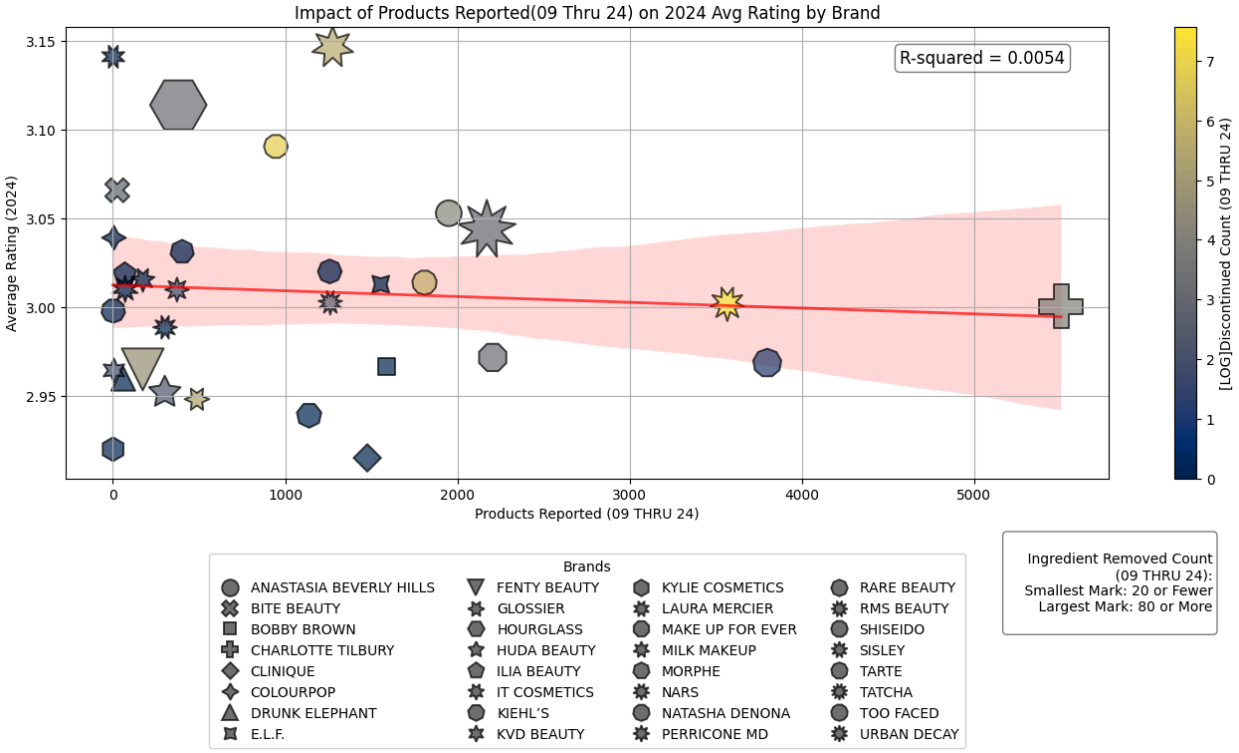
Our merged aggregated dataset allows us to analyze the impact of Products Reported('09 Thru '24), focusing on understanding the relationships between the number of products reported from 2009 through 2024 and their potential effects on two key outcomes: **average rating** and **average price** in 2024. We explore whether trends in product discontinuation or ingredient removal influence brand performance in terms of these outcomes.

1. **Impact on Average Rating**:
    - A linear regression was conducted to analyze the relationship between the number of products reported and the average rating in 2024. The resulting R-squared value for this model was **0.0054**, indicating that only 0.54% of the variability in the 2024 average rating can be explained by the number of products reported between 2009 and 2024.
    - This low R-squared value suggests that there may be other significant factors influencing average ratings besides the number of products reported. Despite the inclusion of markers for discontinued products and ingredient removals, these factors do not appear to have a substantial impact on the overall ratings.
2. **Impact on Average Price**:
    - A second regression analysis was conducted to examine the relationship between the number of products reported and the average price in 2024. The R-squared value for this analysis was **0.0032**, indicating that only 0.32% of the variance in average price can be attributed to the reported product count during this time period.
    - Similar to the rating analysis, the low R-squared value for average price suggests that factors beyond product reporting, discontinuations, or ingredient removals likely have a much stronger influence on price fluctuations.

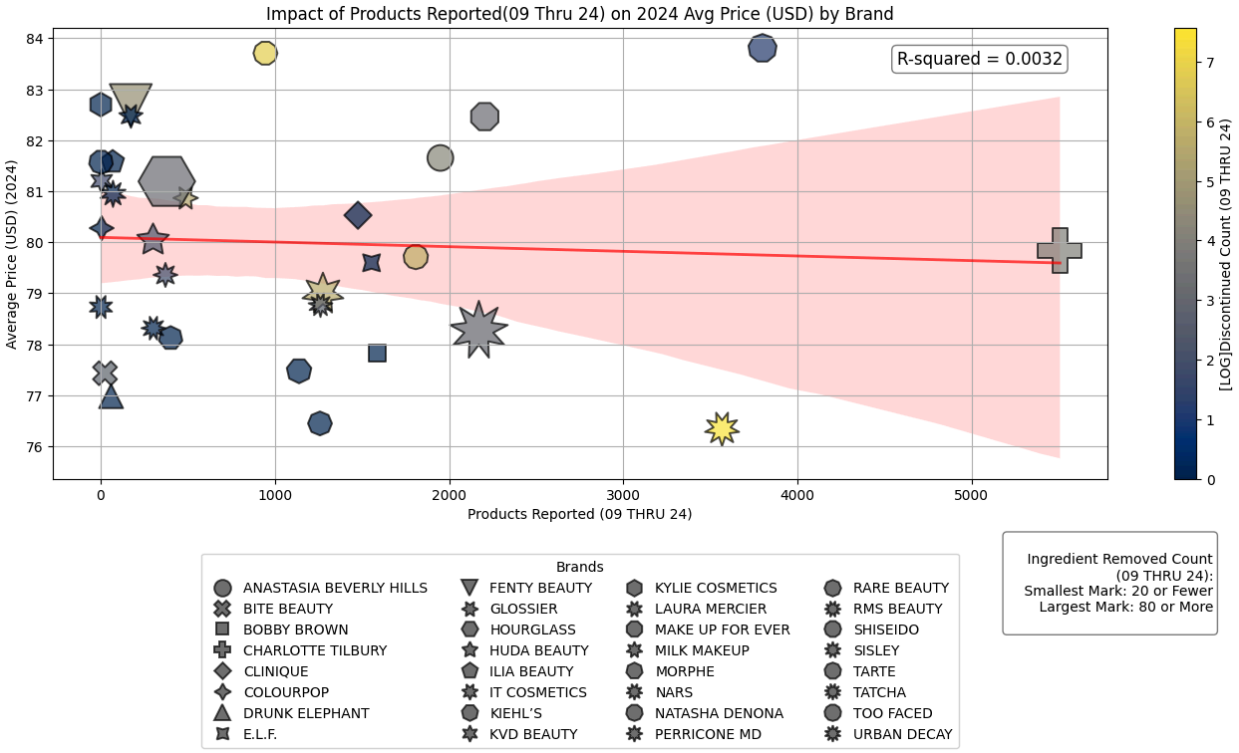
Visualizations

Visualizing Impact of Products Reported(09 Thru 24)

Two comprehensive scatter plots were created to visualize these relationships. Each plot includes a regression line to highlight the linear trend between the number of products reported and the target variable (average rating or price), along with a variety of markers indicating brand-specific variations.



1. **Average Rating Visualization**:
    - This plot displays how each brand's average rating in 2024 is associated with the number of products they reported from 2009 to 2024.
    - Marker size corresponds to the number of ingredient removals, while color intensity (logarithmic scale) reflects the count of discontinued products.
    - Despite clear variations between brands, the weak R-squared value (0.0054) suggests that product-related actions have a negligible impact on ratings.



2. **Average Price Visualization**:
    - Similar to the rating plot, this visualization examines the relationship between reported product counts and the 2024 average price.
    - Brand-level distinctions are represented through different marker styles, and color intensity again indicates discontinued products, with larger marker sizes representing higher ingredient removal counts.
    - The small R-squared value of 0.0032 is even lower than in our previous case, indicating a weaker explanatory power of product reporting on pricing. This reinforces that the relationship between the number of products reported and pricing is less significant compared to the insights gained from the average rating.

Both visualizations offer valuable insights into the marginal influence of product-related actions on brand performance, helping to illuminate the role of public health initiatives and public sentiment in shaping consumer perceptions.

Statement of Work

The project was characterized by strong collaboration, featuring regular check-ins and team discussions facilitated through videoconferencing for real-time engagement. This enabled members to review progress, troubleshoot issues, and provide equal input on the notebook code throughout the project. Each participant played a vital role in identifying key analysis areas, interpreting findings, and designing visualizations. We upheld an open feedback loop, encouraging collective discussions on exploratory results and potential next steps. Below is an

outline of individual contributions, emphasizing the reciprocal input and feedback among all team members.

Cedric Lambert: Contributed to the oversight of the data acquisition process by providing input on the collection, cleaning, and manipulation of product reporting data spanning 2009 to 2024. Conducted linear regression analyses, yielding R-squared values for both average rating and average price. Developed scatter plots to visualize the relationships between product actions and brand performance, highlighting key metrics. He also played a significant role in drafting the analysis as well as the accompanying code notebook, actively incorporating collaborative feedback from team members throughout the process.

References

1. Lignopure. (2024). Importance of transparency in the cosmetic industry. Lignopure. https://www.lignopure.com/importance-of-transparency-in-the-cosmetic-industry/#:~:text=Certification%20and%20Labels:%20Various%20certifications%20and%20labels,product's%20composition%2C%20manufacturing%20processes%2C%20or%20ethical%20considerations.
2. Cosmetics Business. (2024). 68% of consumers seek clean beauty information and Skin Match Technology's Beautytech helps with that. Cosmetics Business. https://cosmeticsbusiness.com/68-of-consumers-seek-clean-beauty-information-and#:~:text=Research%20shows%20that%2060%25%20of,during%20and%20after%20the%20pandemic.
3. Cleanhub. (2023). Clean Beauty Survey 2023. Cleanhub. https://www.cleanhub.com/clean-beauty-survey