Optimizing Cosmetic Formulations Using Product Reformulation Trends

Olivia Folly-Gah
December 6, 2024

This semester, my project, *Product Reformulation Trends: Optimizing Cosmetic Formulations*, centered on analyzing the reformulation of cosmetic products to uncover trends shaped by safety concerns, regulatory changes, and evolving consumer preferences. The objective was to provide actionable insights into how hazardous ingredients influence product updates and discontinuations, thereby supporting the development of safer and more innovative cosmetic formulations. Using data from the California Safe Cosmetics Program (CSCP), I focused on key features such as chemical hazards, reformulation timelines, and product categories to identify meaningful patterns.

Throughout the semester, I laid the groundwork for this analysis by preparing the dataset and conducting exploratory data analysis (EDA). While I did not implement the Naive Bayes Classifier (NBC) this semester, I plan to use it next semester for predicting reformulation likelihood, utilizing the expanded dataset and refined features. This report highlights the findings, challenges, and next steps for the project.

My initial focus was on cleaning and preparing the dataset. I standardized chemical names using CAS Numbers, which resolved inconsistencies and ensured a uniform dataset. I also engineered new features, such as hazard scores from Proposition 65 data, chemical age, and reformulation counts. These features provided the depth needed for meaningful analysis.

The exploratory data analysis phase revealed trends in reformulation activity across different product categories and timelines. For instance, makeup products showed the highest reformulation rates, influenced by regulatory bans and consumer preferences. Temporal patterns highlighted spikes in reformulations following significant regulatory updates, underscoring the industry's responsiveness to external pressures.

One significant observation from the analysis is the relationship between hazard scores and the likelihood of reformulation. Products containing chemicals with higher hazard scores tend to be reformulated or discontinued at a much higher rate. This trend underscores the pivotal role that safety concerns and regulatory pressures play in driving product updates. Quantifying this relationship highlights the importance of rigorous hazard assessment as a key factor influencing reformulation decisions in the cosmetics industry. It emphasizes the need for manufacturers to prioritize safer ingredient alternatives to align with both regulatory requirements and consumer expectations.

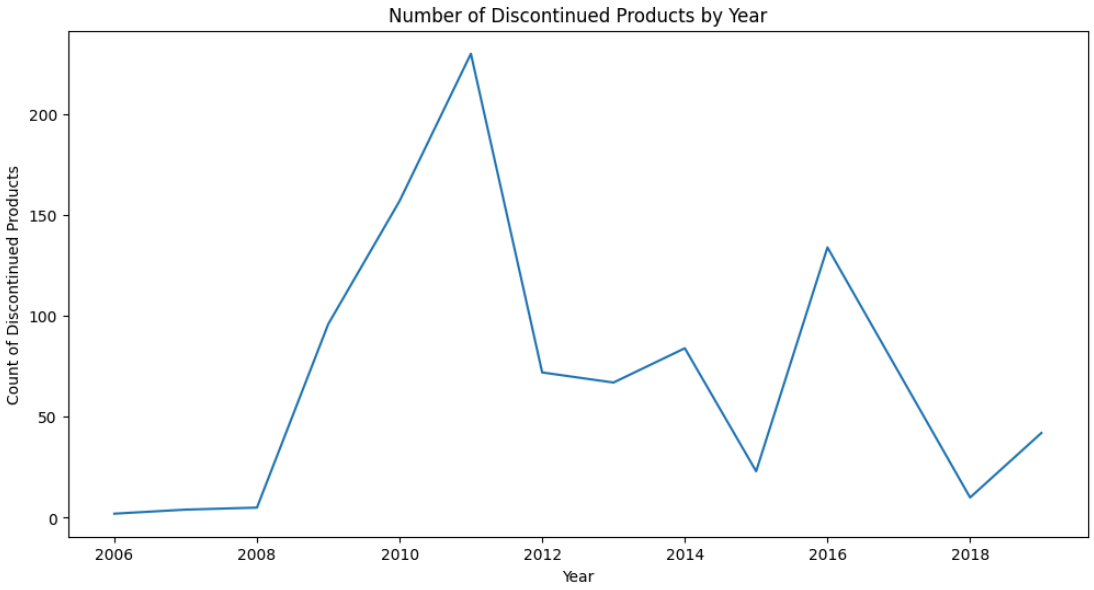

Figure 1: Timeline of Product Discontinuations and Regulatory Updates

This timeline tracks product discontinuations. The analysis reveals that peaks in discontinuation activity often coincide with the implementation of new regulations or updates to existing safety guidelines. For example, an increase in product removals occurred following a major Prop 65 update in 2016. This figure emphasizes how external regulatory forces directly impact industry practices.

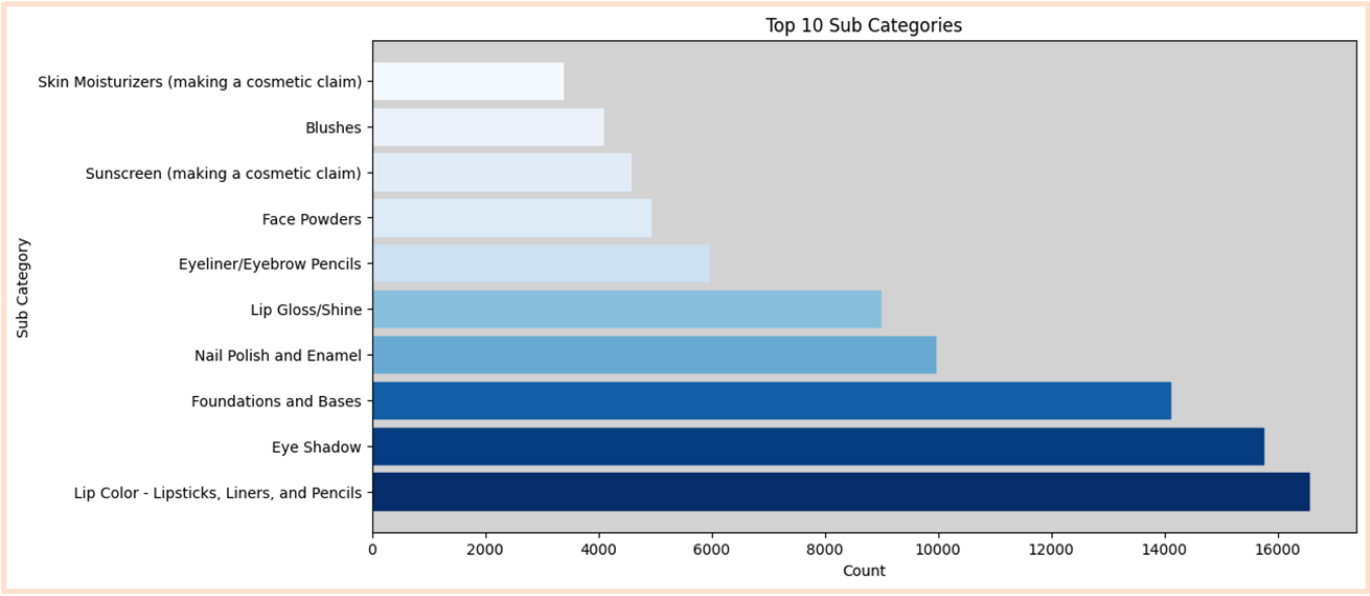

Figure 2: Reformulation Frequency by Product Subcategories

This figure highlights the reformulation frequency across the top ten product subcategories. Makeup products rank highest, followed by hair care and skin care items. The analysis suggests that makeup products face the most reformulations due to a combination of stringent safety regulations and rapidly evolving consumer preferences. This insight is crucial for identifying product categories most susceptible to reformulation trends.

The progress this semester established a solid foundation for the project. Data cleaning and feature engineering resolved inconsistencies and enriched the dataset. EDA provided key insights into reformulation trends, setting the stage for predictive modeling. However, certain tasks remain pending, such as the implementation of the Naive Bayes Classifier and the expansion of the dataset to include formulations tailored to coily hair and melanated skin. These steps will be prioritized next semester.

One of the main challenges was inconsistent naming conventions for chemicals and companies, which I addressed by standardizing data using CAS Numbers. Another limitation was the dataset's lack of diversity, particularly in formulations relevant to coily hair and melanated skin. While this issue remains unresolved, I plan to address it next semester by incorporating data from external sources.

In addition, I identified the need for more numerical features, such as ingredient percentages and pH levels, to enhance model accuracy. These enhancements will enable the Naive Bayes Classifier to provide more reliable predictions in the next phase of the project.

While I did not use the Naive Bayes Classifier this semester, I plan to implement it next semester. The technique will allow me to classify products based on reformulation likelihood using features such as hazard scores, ingredient presence, and reformulation history. Combined with an expanded dataset and additional numerical features, this approach will provide deeper insights into reformulation drivers.

Next semester, I plan to expand the dataset to include formulations tailored to coily hair and melanated skin. This will address a critical gap and ensure the analysis reflects diverse consumer needs. I will also incorporate new numerical features, such as pH levels and ingredient percentages, to enhance the dataset's utility.

To complement these efforts, I plan to implement the Naive Bayes Classifier to predict reformulation likelihood and develop an interactive Tableau dashboard to visualize trends dynamically. These tools will make the analysis more accessible and actionable for stakeholders.

This semester laid the groundwork for understanding product reformulation trends in the cosmetics industry. By addressing challenges in data standardization and leveraging exploratory data analysis, I gained valuable insights into how regulatory changes and ingredient hazards shape product updates.

Looking ahead, I aim to build on these findings by expanding the dataset and applying machine learning techniques, such as the Naive Bayes Classifier, to predict reformulation needs. This project represents an important step toward bridging data science and cosmetic chemistry, with the goal of advancing both industry practices and consumer safety.