Identification of Hazardous Ingredients in Cosmetic Products for Reformulation

Olivia Folly-Gah

11/1/2024

Introduction

The goal of my project is to identify harmful chemicals in cosmetics that may contribute to cancer, birth defects, or reproductive harm. By analyzing data from the California Department of Public Health (CDPH), I aim to highlight frequently used hazardous substances in cosmetic products and generate insights that can support safer reformulation practices across the industry. This report outlines the project's progress so far, the challenges encountered, and the planned next steps to further advance this analysis.

Summary

During this period, the primary tasks completed include data preprocessing, analysis of product discontinuation trends, and initial exploration of hazardous chemical concentrations across different product subcategories. I have carefully processed the dataset to handle inconsistencies and prepare it for in-depth analysis, allowing for accurate identification of products containing hazardous chemicals. By examining historical trends, I have identified years with significant spikes in product discontinuations, which could correlate with regulatory changes or shifts in consumer demand for safer products. Additionally, I explored the prevalence of hazardous chemicals within the top 10 cosmetic subcategories, which highlights that makeup products contain the highest levels of such ingredients. These findings provide a valuable basis for further analysis of safety trends and reformulation needs in the cosmetics industry.

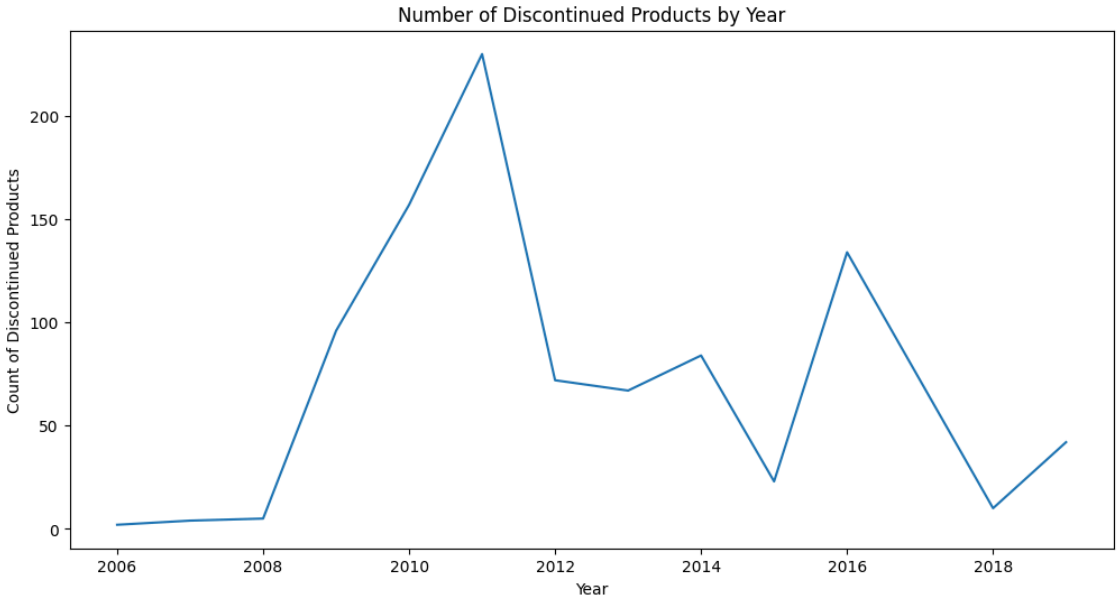

Figure 1: The first visualization illustrates the trend of product discontinuations by year.

This chart reveals shifts in the rate of discontinued products, likely in response to regulatory changes and consumer awareness surrounding cosmetic safety. In particular, the data shows a gradual increase in discontinued products between 2006 and 2008, suggesting a growing awareness of cosmetic safety issues during that period. The discontinuation rate peaks in 2010, which could be due to new regulations or increased scrutiny of hazardous substances in cosmetics. After 2010, the rate of discontinuations declined, indicating potential stabilization as manufacturers adapted to regulatory requirements. However, there is a resurgence in 2016, suggesting renewed industry scrutiny over existing products and chemicals.

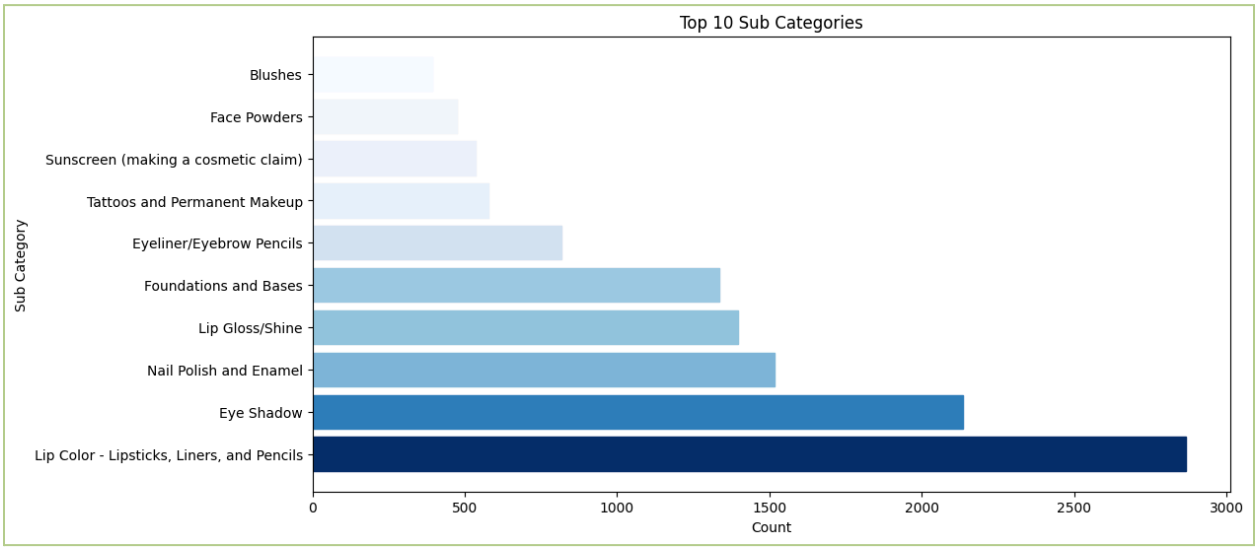

Figure 2: The second visualization ranks the top 10 cosmetic subcategories by the presence of hazardous chemicals.

Makeup, among other categories, shows the highest concentration of hazardous ingredients, indicating a need for additional reformulation efforts within this category. This preliminary analysis provides insights into which subcategories require the most attention in terms of consumer safety and regulatory compliance.

Progress and Milestones

In the past two weeks, I successfully completed the initial data preprocessing tasks necessary for a clean and structured dataset. These tasks involved resolving inconsistencies within the data and preparing it for robust analysis of chemical hazards and product discontinuation trends. I conducted a preliminary examination of the discontinued products, identifying significant spikes in discontinuations that correspond with historical shifts in safety regulations. Additionally, I analyzed hazardous chemical presence within different product subcategories, focusing on makeup and other frequently used products. These analyses contribute to the foundational understanding of current cosmetic safety practices and reformulation priorities.

Problem-Solving and Challenges

One of the significant challenges encountered was managing and interpreting discontinuation trends within a large, complex dataset. This required extensive preprocessing to ensure accurate chronological alignment of product entries, especially given the fluctuations in reporting frequencies and varying product lifecycle timelines. To address this, I utilized statistical techniques to standardize date formats and clean up any redundancies within the data. Another challenge was determining how to effectively categorize products with overlapping classifications. By using clear criteria to distinguish between subcategories, I ensured that data visualizations and subsequent analyses were consistent and meaningful. The solutions implemented have enabled a clearer view of discontinuation trends and a reliable categorization of hazardous chemical concentrations within the top subcategories.

Technical Depth and Accuracy

The main technical approach involved creating detailed visualizations to represent trends and subcategory rankings. For Figure 1, which displays product discontinuation trends by year, I used time series analysis techniques to identify and interpret shifts in the data. This allowed me to connect regulatory milestones with peaks in product discontinuations, providing insights into industry responses to increased scrutiny on cosmetic ingredients. Figure 2 ranks the top 10 subcategories by hazardous chemical concentration, highlighting that makeup products contain the highest levels of such ingredients. This analysis emphasizes the importance of targeting makeup formulations for safer alternatives. Further, I performed statistical analyses to confirm the accuracy of my subcategory rankings, ensuring the integrity of my findings.

Future Plans and Goals

In the next two weeks, I plan to expand upon the initial findings by analyzing specific hazardous chemicals that have been most frequently discontinued and examining patterns in consumer demand. By linking chemical hazards to historical regulatory changes, I aim to assess the long-term impact of these substances on cosmetic formulations and industry safety practices. Before applying any modeling, I will address the imbalanced nature of my dataset to ensure that the analyses yield reliable and actionable insights. Additionally, I will conduct more granular analyses of reformulation strategies within high-risk subcategories, focusing on makeup products to identify ingredients that may require reformulation. I will also explore methods for segmenting the dataset further to provide more precise recommendations for safe product development within the cosmetics industry.