

Kaydence Lin | lin.kay@husky.neu.edu

Andrea Keiper | keiper.a@husky.neu.edu

DS2001

April 13, 2023

Chemicals in Cosmetics

Problem statement/background

There are ingredients and chemicals that are found in cosmetic products that can be harmful and can cause long-term health issues, such as cancer and allergies, that most consumers are often unaware of. This can affect all people of all ages across the world. Manufacturers, packers, and/or distributors named on product labels must give the California Safe Cosmetics Program a list of all cosmetic products that contain any ingredients known or suspected to cause cancer, birth defects, or other developmental or reproductive harm. It includes data such as product name, brand name, cosmetic categories, chemical name, dates reported, and discontinued date.

Introduction and Description of the Data

We chose this data set due to its importance in terms of public health. The large majority of people use a sum of cosmetic products in their daily life, including both of us. When we found this data set published by the California Safe Cosmetics Program (CSCP) in the California Department of Public Health (CDPH), we knew it was a trustworthy data set that could offer information about what chemicals are included in all different types of cosmetic products from many different companies. Many potentially harmful chemicals are included in products that we use on or in our bodies, and it is crucial to understand what these chemicals are and what they have the potential of doing to us. We believe that the majority of consumers using cosmetic

products would be interested in this problem. With a better awareness of the chemicals in our products and the threats they pose, every consumer will be able to make safer purchases for their personal health and well being.

Methods

The first function we created was the `read_file` method which reads the data from each row in the file, filters out any rows that include a chemical that was removed from the product, and adds all other lines to a list. Our second function was `get_stats` which first read in the data of a specific column, then we used the `Counter()`₂ method to get the overall count of each unique item within the column. We used this function to count the number of products each brand had on the list, and the number of products each cosmetic category had on the list. We then used the `most_common()`₁ function to get the top 5 counts of a category within the column in order to make visuals. The `bar_chart` method returns a bar chart using the name and amount of something, which we used to report the number of products per brand and the number of products per cosmetic category. The `pie_chart` method returns a pie chart using the name and amount of something, which we used to report the percentage makeup of each of the top chemicals included on the list.

Results, Conclusions and Future Work

With our work, we wanted to create visuals that accurately represented the data and could potentially help consumers. Figure 1 shows us that the most frequently repeated brands on the list were Sephora, NYX, bareMinerals, Revlon, and Tarte. The figure also shows the number of items included on this list for each of these brands. This does not necessarily mean that Sephora products are inherently worse than all other brands on this list, but it does mean that Sephora and the following brands are certainly ones to look out for when purchasing cosmetic items, as they

may include potentially harmful products. Sephora has accumulated over 5,000 products that contain potentially harmful chemicals, followed by NYX and bareMinerals with over 3,000 products each.

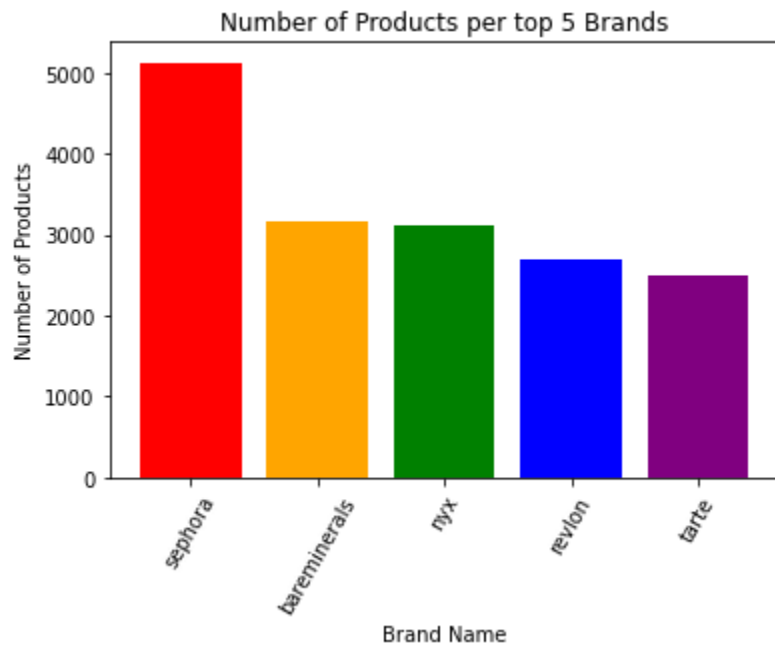


Figure 1. Number of products included on the list for each of the 5 most frequently included brands.

Figure 2 displays the number of products included on the list for each of the five most frequently included cosmetic categories. Makeup products make up an alarming 66.5% of products on the list (74,270 products), followed far behind by nail products, skin care products, sun-related products, and bath products. Similar to figure 1, this should display to consumers and cosmetic product users that they must be cognizant of the dangers of the products they may be purchasing. This could make consumers make safer purchases when choosing products from these five categories.

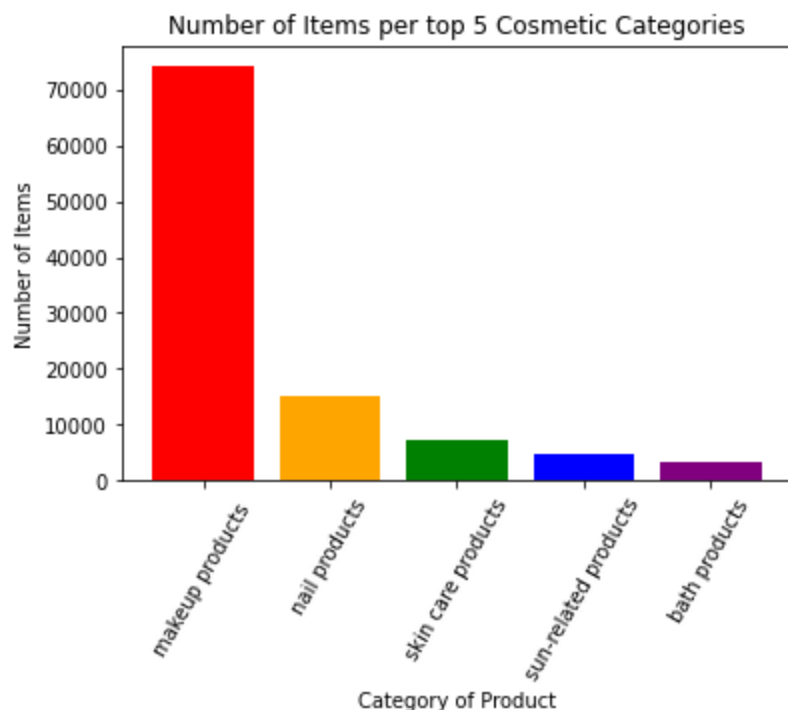


Figure 2. Number of products included on the list for each of the 5 most frequently included cosmetic categories

Figure 3 displays which five chemicals are the most frequently reported from our data set. First is titanium dioxide at 91.6%, followed far behind by silica, retinol/retinyl esters, mica, and butylated hydroxyanisole. Titanium dioxide can irritate the eyes, nose and throat, potential carcinogen and cause of reproductive damage. Butylated hydroxyanisole is a potential carcinogen and affects hormone behavior. Retinol/retinyl esters can cause skin irritations, such as redness, dryness and flaking of the skin. Silica can cause silicosis: a lung condition that causes inflammation in the lining of the lung. Mica can be dangerous if it is inhaled, it contains particles that can cause lung scarring. While the other two figures are very important, this one may be the most crucial for consumers to see and take note of. Knowing the most frequently used chemicals in cosmetic products will allow consumers to avoid potential harm by reading the ingredient lists on their products and spotting these dangers. While it is helpful to know what brands and

categories are frequent offenders, consumers should always read the ingredients when considering a product, no matter what brand or category. By offering the data from figure 3, the consumer is able to take into account five specific chemicals that they are likely to run into that are known to cause potential harm.

Most Common Chemicals Used in Cosmetics

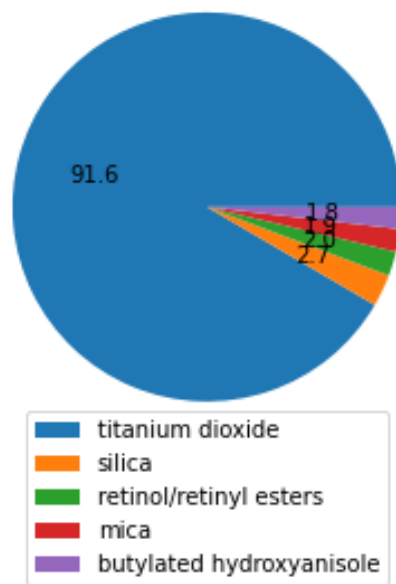


Figure 3. The most common chemicals included on the list

We believe that our visuals are the strengths of our project, as they offer important and helpful information given the data set and the goal of offering knowledge to consumers. A weakness of the data set is that it includes chemicals that have been proven to be not harmful in trace amounts or in amounts approved by the FDA and other scientists. Many products are included on the list that contain chemicals that have been tested to be safe in small amounts, but potentially dangerous in different forms or amounts that would never be seen for use in cosmetic products. Therefore in our visuals, the brands, categories, and even chemicals themselves are not necessarily an accurate representation of what dangers are being presented to consumers and by whom. At the same time, it is crucial that consumers are aware of what they are putting onto

and/or into their bodies, and this data is beneficial in that way. Given more time, we could have done further research and filtered out chemicals that have been proven to not cause true harm in the small amounts that are being used in the project. This would take time because we would have to confirm that each product on the list only contains the minimum amount of the chemical that is considered safe. In the future, we can analyze the data set again, as it is updated constantly, to see if there are any significant changes in brands, products, or ingredients. We can use this data and compare it to laws and regulations that have been created regarding chemicals in cosmetics and their implications.

Works Cited

1. “Count Elements in a List with Collections.counter in Python.” *Count Elements in a List with Collections.Counter in Python*, <https://note.nkmk.me/en/python-collections-counter/>.
2. Pozo Ramos, Leodanis. “Python's Counter: The Pythonic Way to Count Objects.” *Real Python*, Real Python, 25 Jan. 2023, <https://realpython.com/python-counter/>.