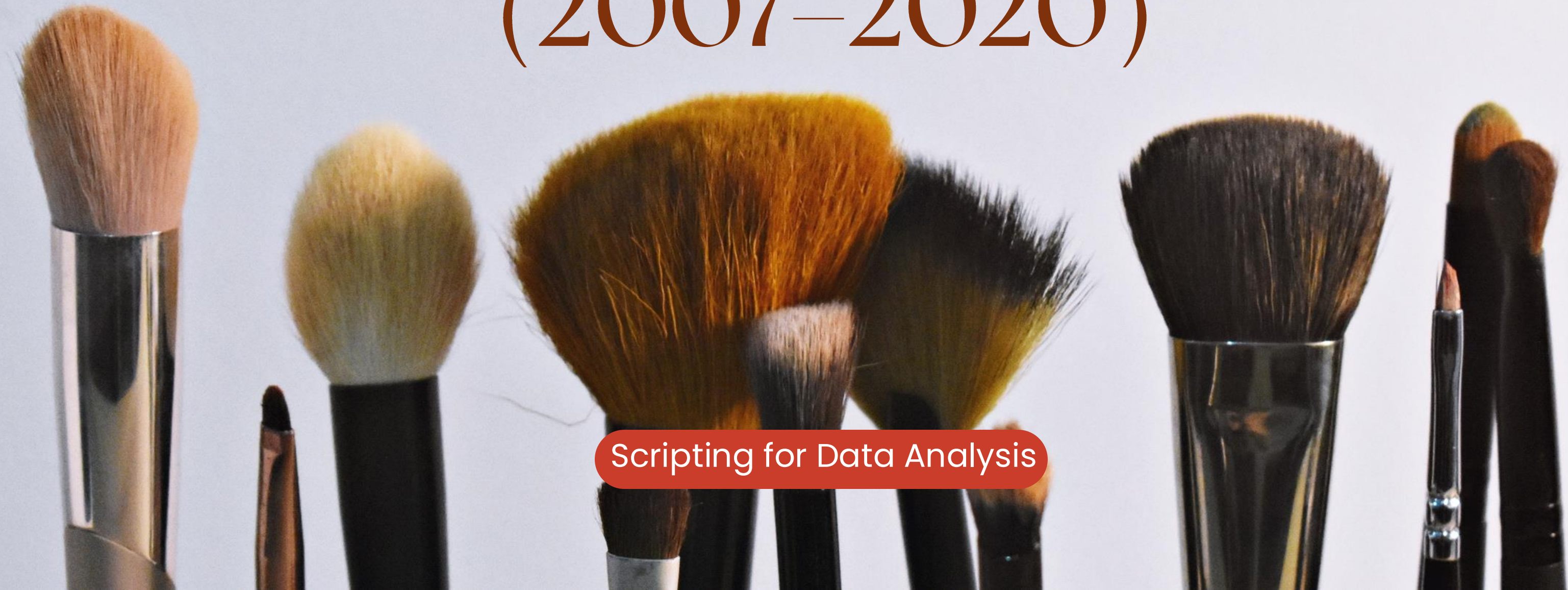


ANALYZING HAZARDOUS COSMETIC CHEMICAL DISCLOSURES IN CALIFORNIA (2007–2020)



Scripting for Data Analysis

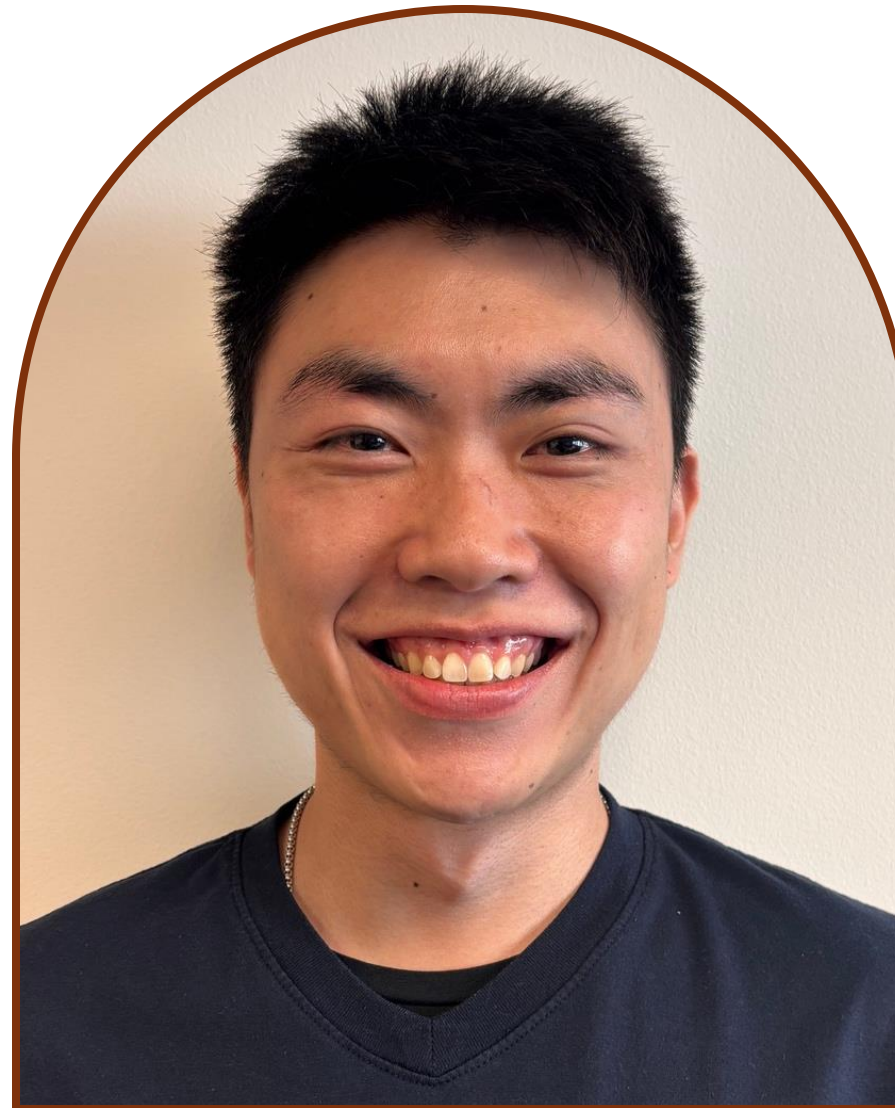
Our Team



Ashish Chaudhary



Mian Hamid



Kasem Chaisathit.



Prerna Bharti

Introduction to our Project

- The California Safe Cosmetics Program requires companies to report chemicals linked to cancer, birth defects, or reproductive harm.
- This project analyzes over 114,000 ingredient disclosure records.
- Goal: understand hazardous chemical patterns, company behavior, chemical usage, and discontinuation trends.



Dataset Overview

Source: California Safe Cosmetics Program Public Dataset

Time Range: 2007–2020

Original Size: 114,635 rows × 22 columns

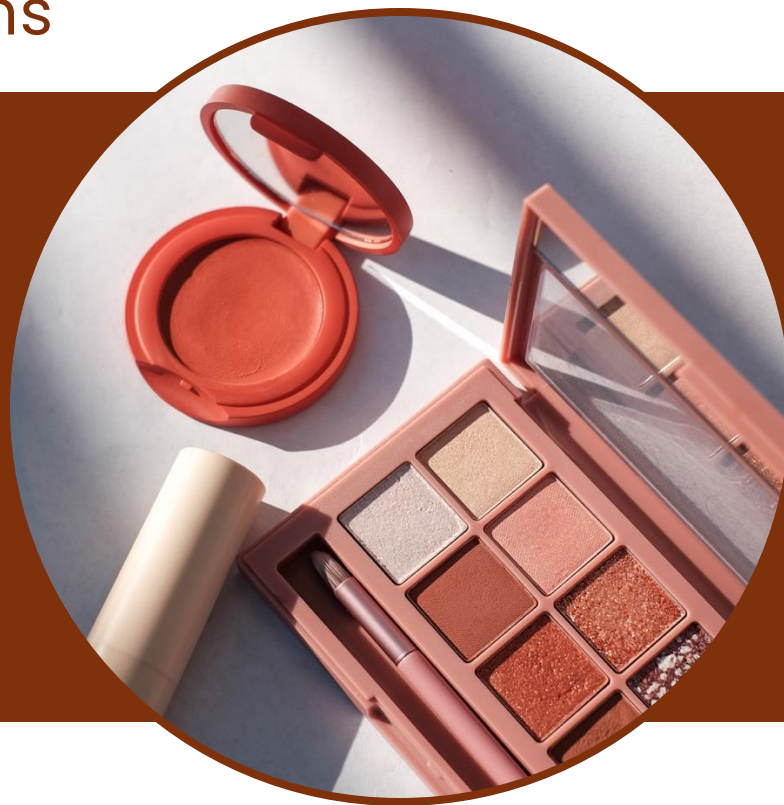
Unique observations:

- Cosmetics products
- Reporting companies
- Chemicals flagged for health concerns



Why this dataset?

- ***Large, public, policy-relevant dataset***



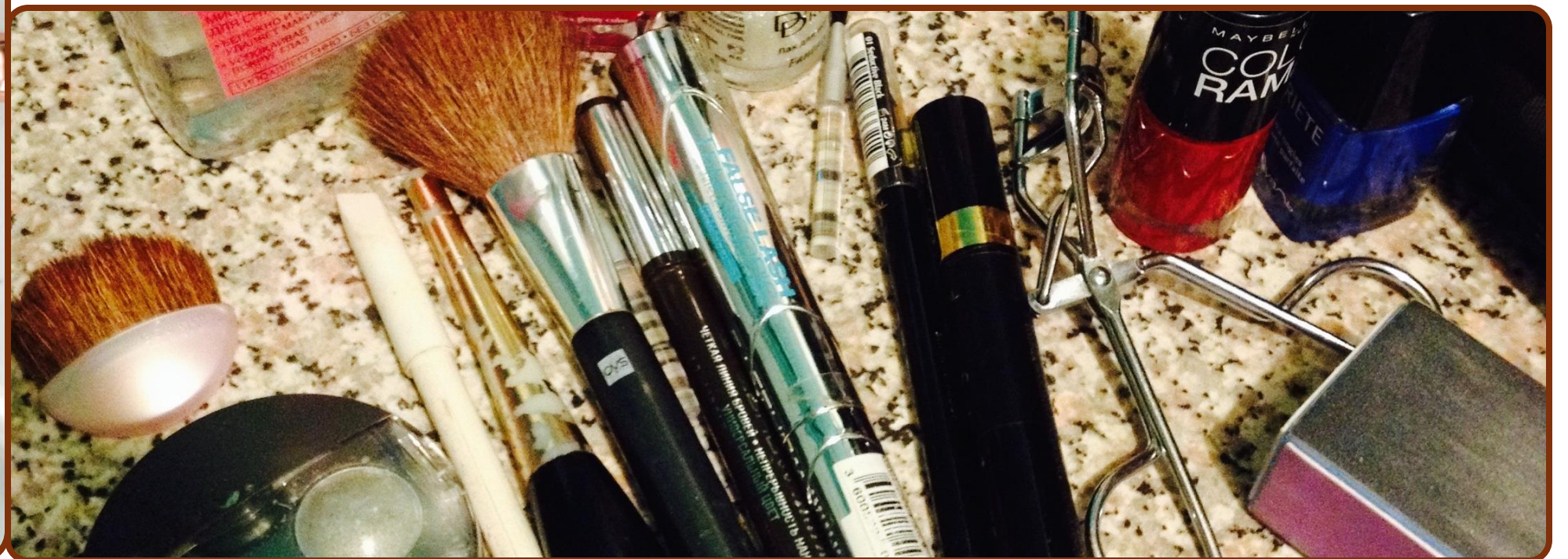
- ***Rich chemical, company, and product information***
- ***Allows multi-dimensional analysis (product → chemical → company)***

Key Columns Used

We cleaned and focused on 13 relevant fields:

- CDPHid (Product ID), ProductName, CompanyId, CompanyName, BrandName, PrimaryCategory, SubCategoryId, SubCategory, ChemicalName, DiscontinuedDate, ChemicalCount, Discontinued

From 22 features to 13!



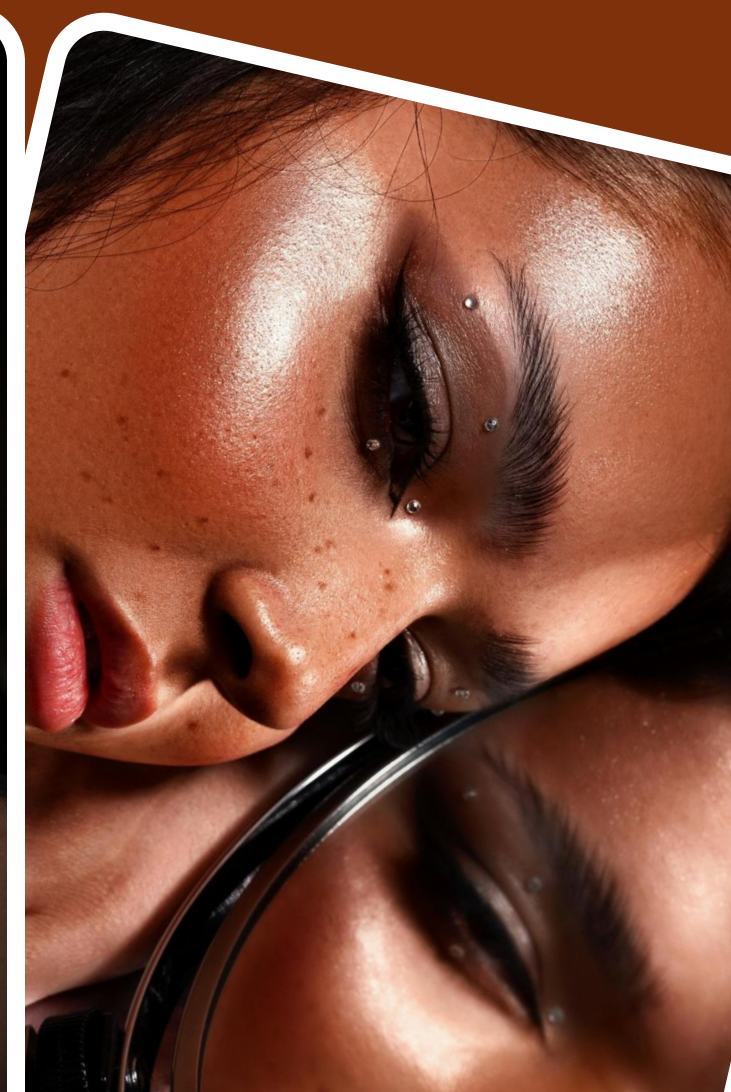
Data Cleaning Summary

Main cleaning steps:

- Removed confidential or unused fields (CSF, timestamps, redundant IDs)
- Fixed missing BrandName rows
- Created a binary Discontinued column
- Deduplicated by Product × Chemical × Company
- Recalculated chemical count per product
- Final cleaned dataset size: 41,762 rows

70,000+

**Duplicate and Missing
Records have been Removed!**



Analysis Overview

This project answers 4 major questions:



What are the most frequently used chemicals?



Which companies use the highest number of chemicals per product on average?

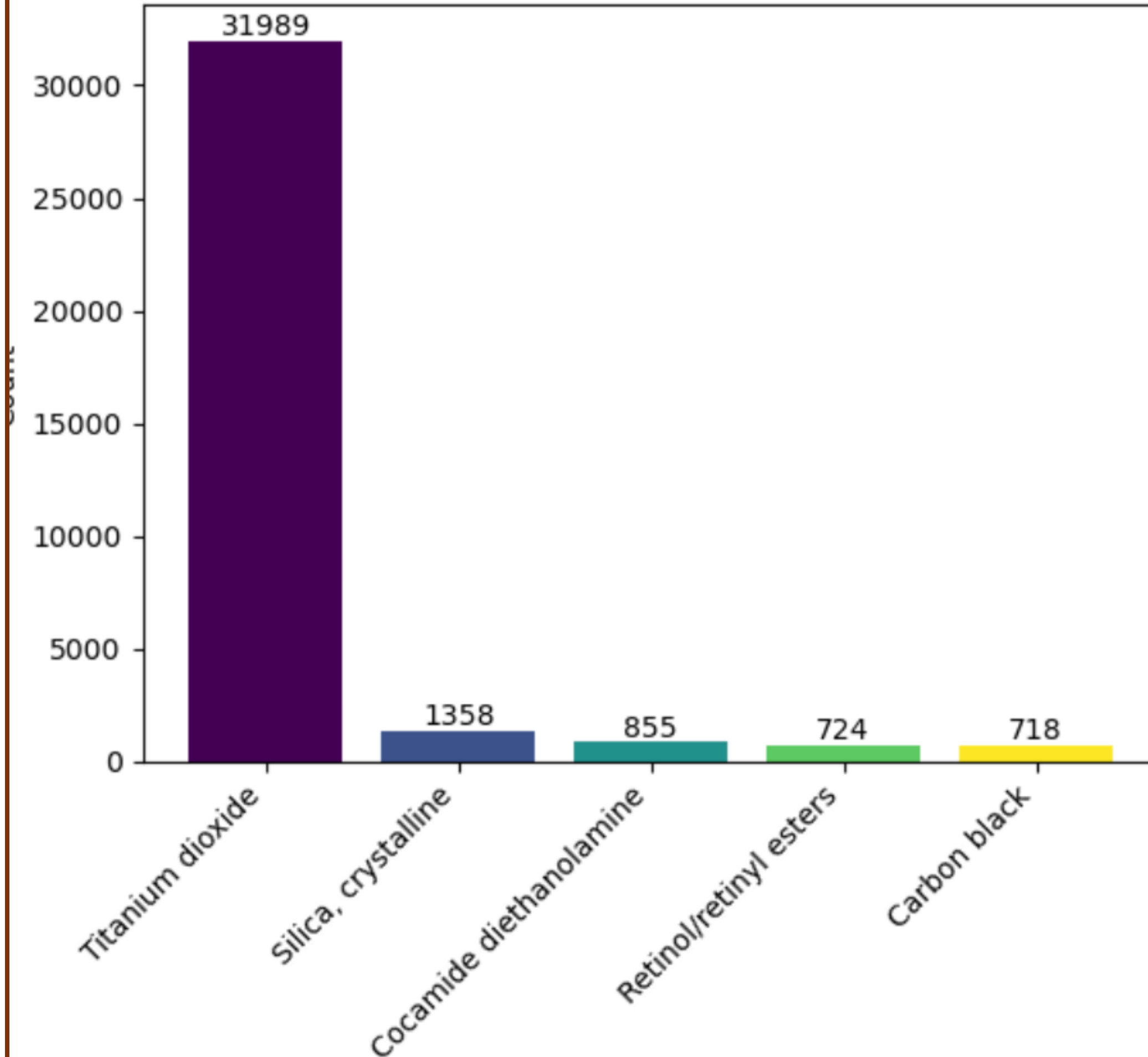


Which product categories carry the heaviest chemical load?



Are products with more chemicals more likely to be discontinued?

Top 5 Chemicals



Q1: Top 5 Most Frequently Used Chemicals

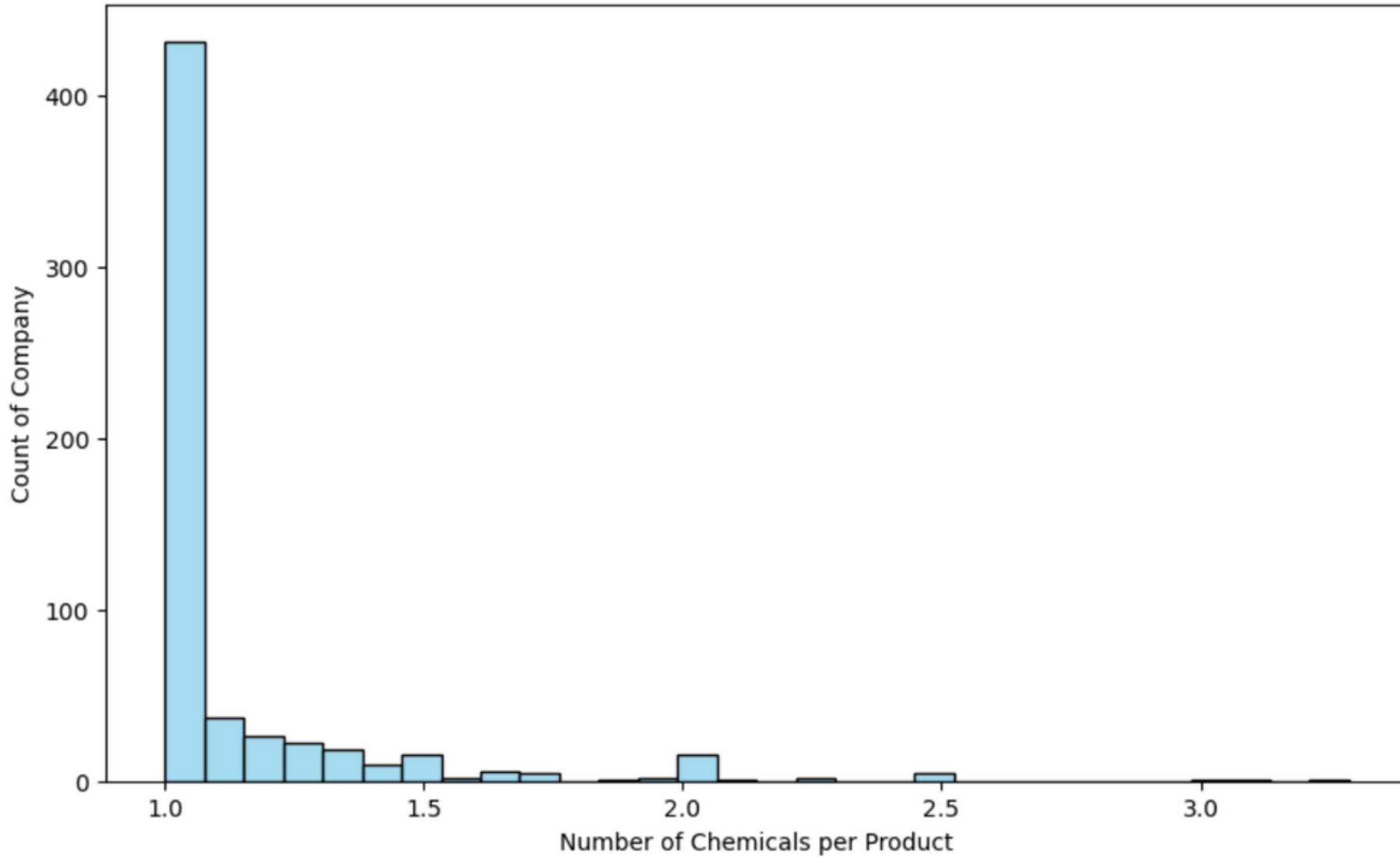
Findings:

- Titanium dioxide dominates usage with 31,989 appearances
- Next highest chemicals include:
 - Silica, crystalline
 - Cocamide diethanolamine
 - Retinol/retinyl esters
 - Carbon black

Interpretation:

- **Titanium dioxide appears across makeup, skincare, and sun-related products**
- **Heavy reliance on a small set of chemicals raises regulatory and safety importance**

Average Chemical Count per Product By Company

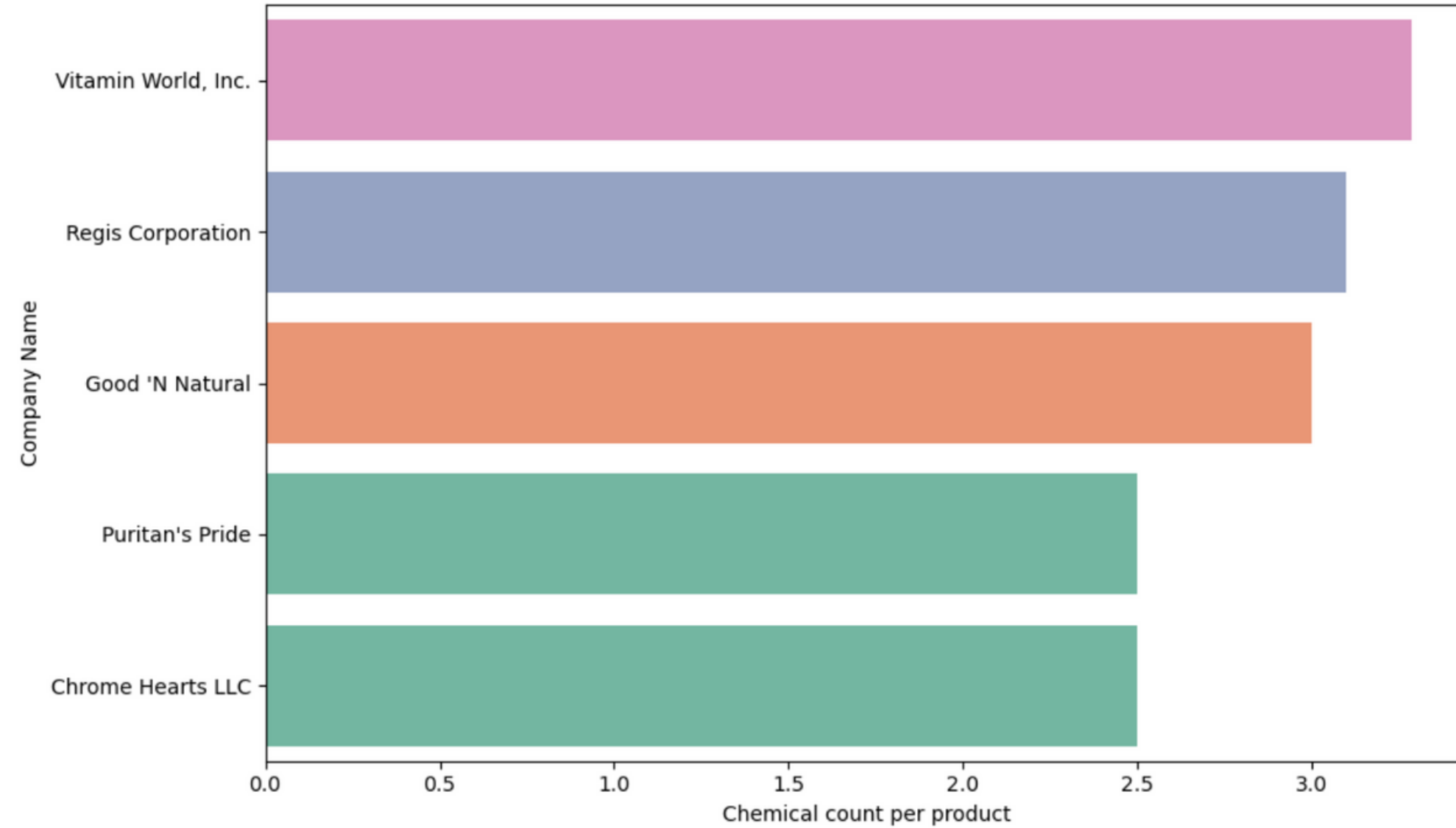


Findings:

- Most companies average 1 chemical per product
- A few companies use 3+ chemicals per product on average:
 - Vitamin World
 - Regis Corporation
 - Good 'N Natural
 - Puritan's Pride
 - Chrome Hearts LLC

Q2: Chemical Count per Product by Company

Top 5 Companies by Average Chemical Count per Product



Interpretation:

- **These outlier companies tend to produce chemically complex formulations**
- **May indicate specialized product lines (e.g., hair treatments, supplements, styling products)**

Q3: Which Categories of Cosmetic Products Contain the Most Chemicals?

Products with the highest number of chemicals

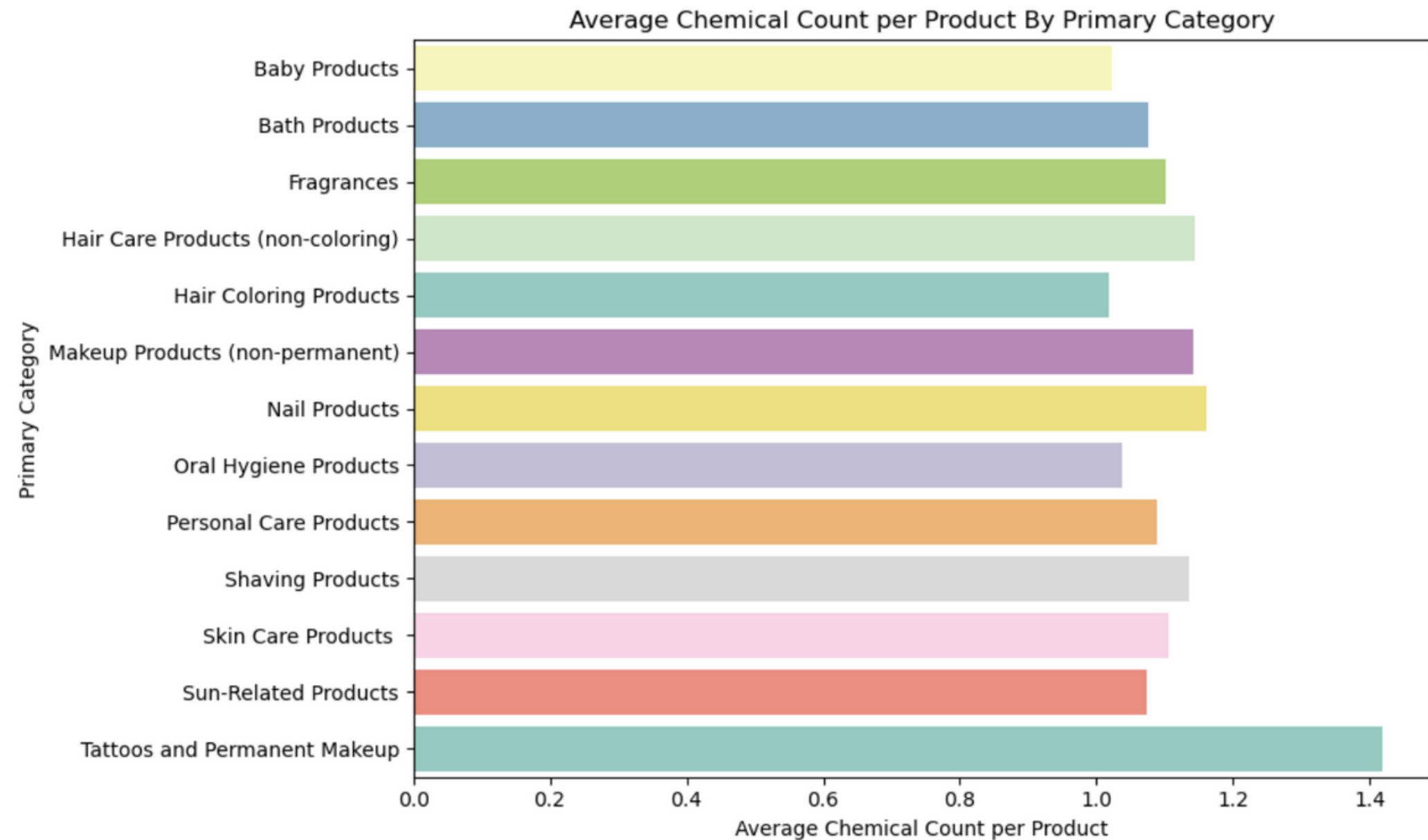
ProductName	BrandName	PrimaryCategory	SubCategory	ChemicalCount
Interrupt Texturizing Clay	Regis Design Line	Hair Care Products (non-coloring)	Hair Styling Products	9
Moisturizing Shampoo	Regis Design Line	Hair Care Products (non-coloring)	Hair Shampoos (making a cosmetic claim)	9

Top categories by chemical density:

1. Tattoos & Permanent Makeup
2. Nail Products
3. Hair Care Products (non-coloring)
4. Makeup Products (non-permanent)

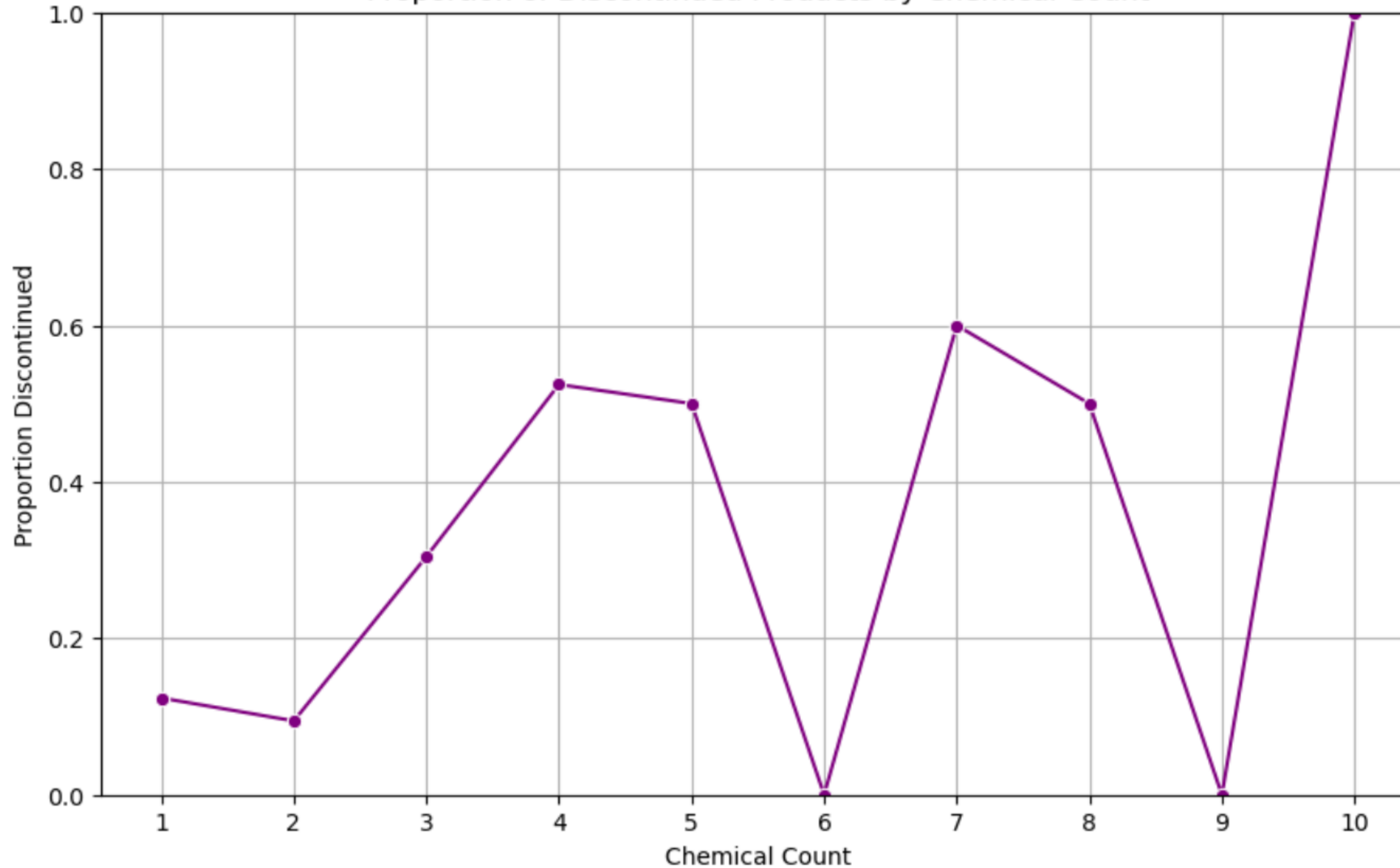
Interpretation:

- **These categories rely more on solvents, preservatives, and fixatives**
- **Higher chemical complexity may relate to durability, stability, or application performance**
- **Baby products, hair coloring and oral hygiene products had the lowest average chemical counts**



Q4: Are Products With More Chemicals More Likely to Be Discontinued?

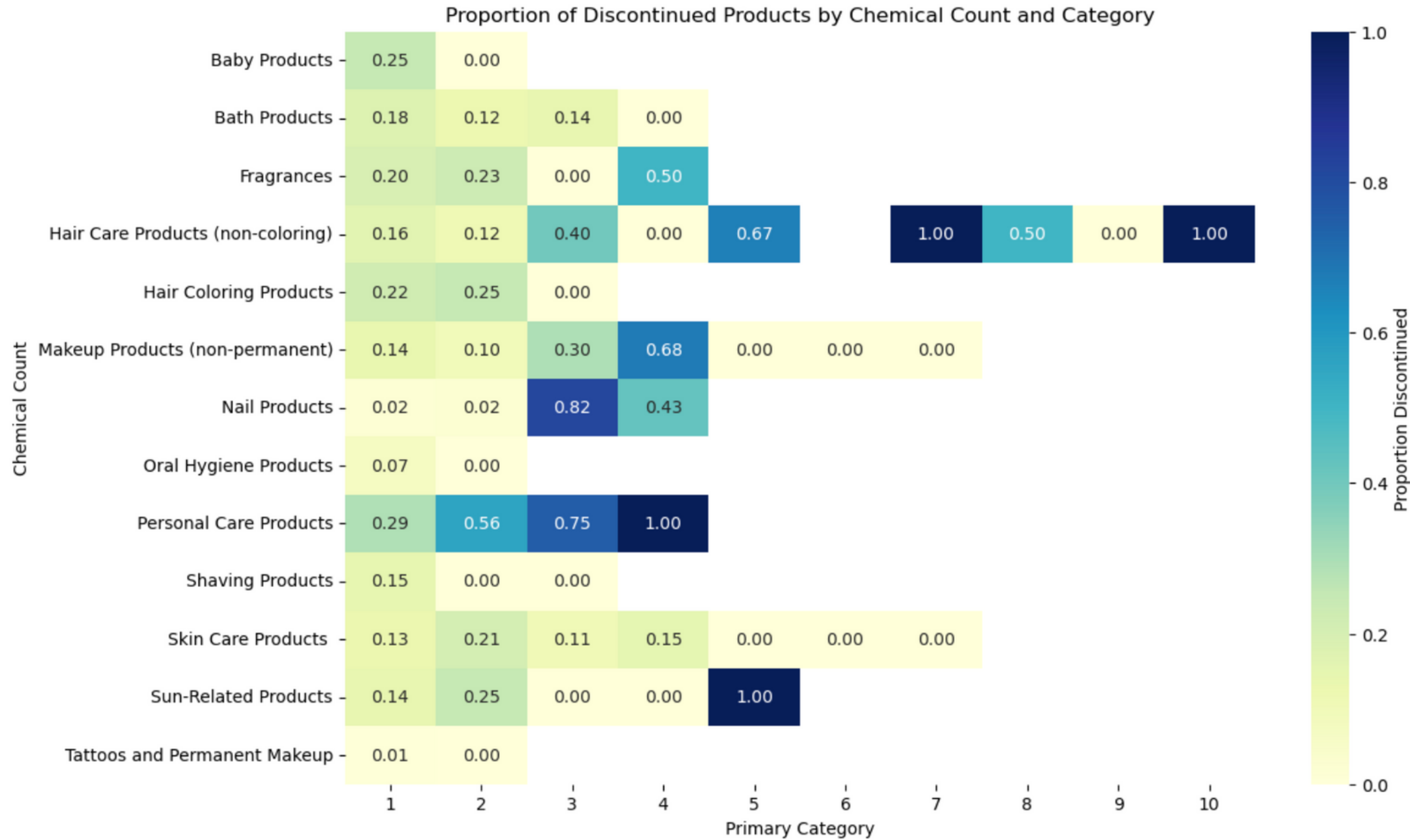
Proportion of Discontinued Products by Chemical Count



Key result:

- Discontinued probability rises sharply as chemical count increases
 - 1–2 chemicals → ~10% discontinued
 - 3–4 chemicals → 30–50% discontinued
 - 7–10 chemicals → some categories reach 100% discontinued (rare but notable)

Q4 (cont.)



Interpretation:

- **Products with higher chemical complexity appear harder to maintain**
- **Could reflect:**
 - **safety recalls**
 - **reformulations**
 - **regulatory compliance issues**
 - **market performance**
- **Heatmap shows categories like Personal Care & Hair Care have especially high discontinuation at higher chemical counts.**

Summary of Findings

Main conclusions:

- A small number of chemicals dominate the industry
- Companies vary widely in chemical usage; a few are consistent heavy users
- Some product categories inherently require more chemicals, e.g., Titanium dioxide is safe on skin in cosmetics, but can be harmful if inhaled in fine powder form.
- More chemicals → higher chance of product discontinuation
- Suggests that complexity increases regulatory or performance risk



 IST652 SECTION M001

THANK YOU

QUESTIONS?