

What's In My Makeup Bag?

Galvanize DSI Capstone 1
Chelsea Ramos



Project Description & Motivation

- Analyzed records of cosmetic products with known/suspected harmful chemicals
- Used data from a CSV file with over 110,000 records
 - Reported to the California Safe Cosmetics Program from 2009-2020, updated daily
 - Had over 600 companies, 120 chemicals
- Was motivated as a cosmetics user to find a safer company





Data Wrangling

- Loaded csv into a Pandas DataFrame
- Focused on less than half of the 22 columns, most importantly:
 - CDPHId (identifies unique products)
 - CompanyName
 - *DiscontinuedDate* (to classify discontinued products)
 - *ChemicalDateRemoved* (to classify reformulated products)
- Classified products as discontinued/reformulated with date columns
- Cleaned categorical columns

CompanyName	BrandName		Cleaned_CompanyName	Cleaned_BrandName
Fresh, Inc.	Fresh		fresh inc	fresh
Fresh Inc.	Fresh		fresh inc	fresh
Fresh Inc	Fresh		fresh inc	fresh

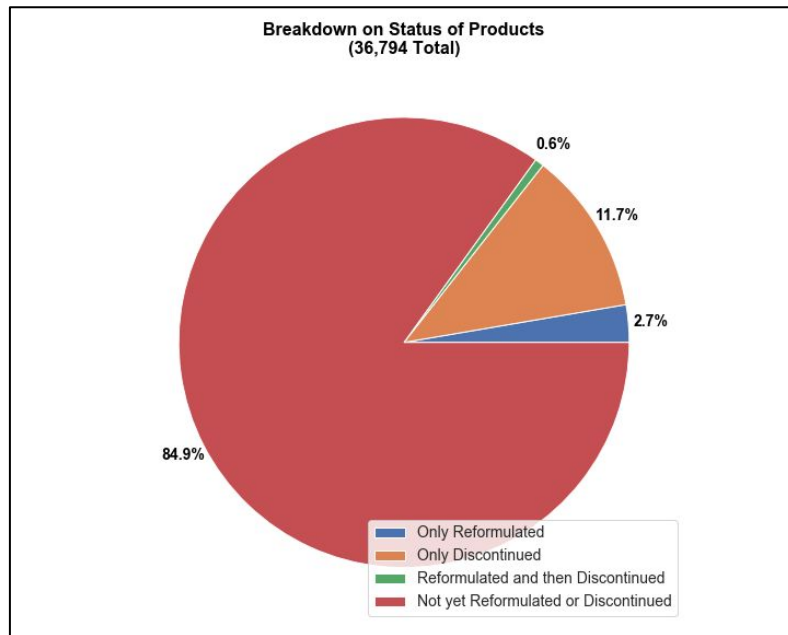


EDA



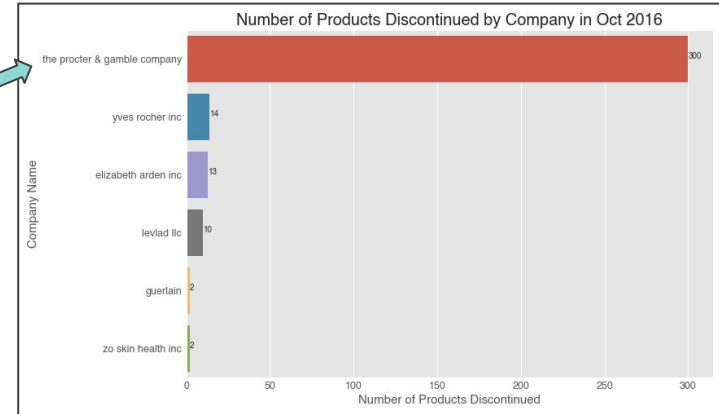
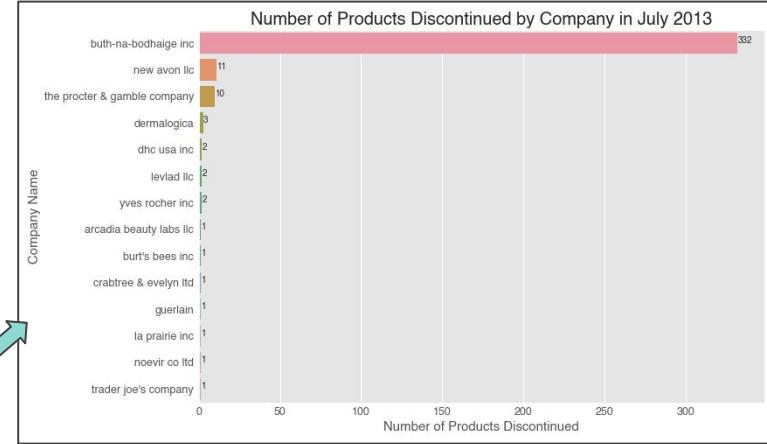
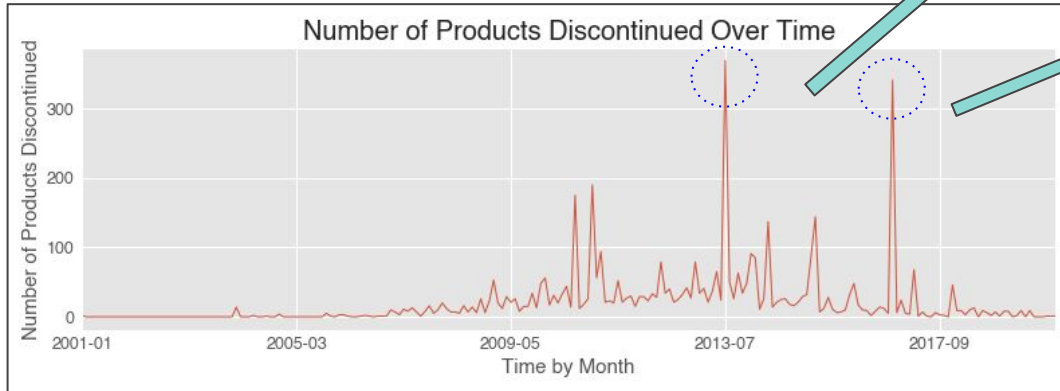
Status of Reported Products

- Nearly 37,000 products reported
- Around 31,500 not yet reformulated/discontinued
- 4,500 were discontinued and 1,000 were only reformulated



Product Discontinuations

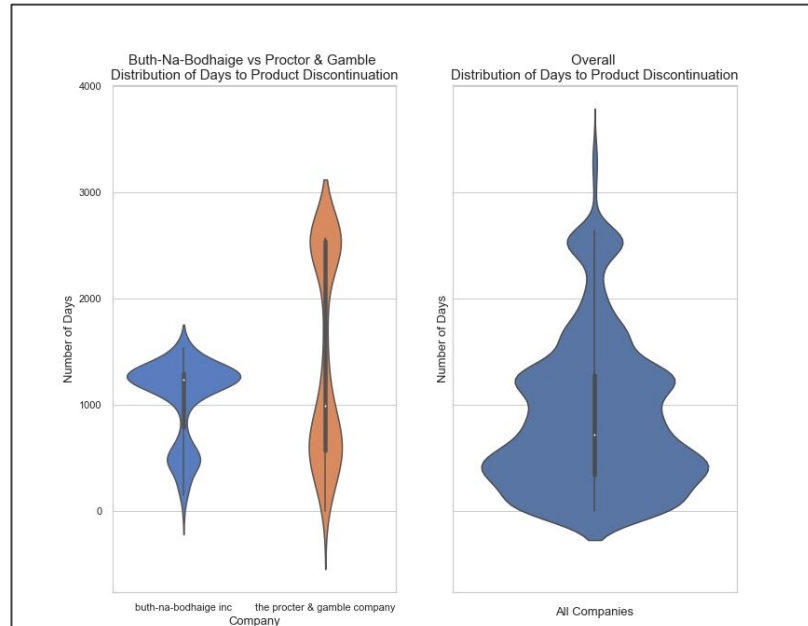
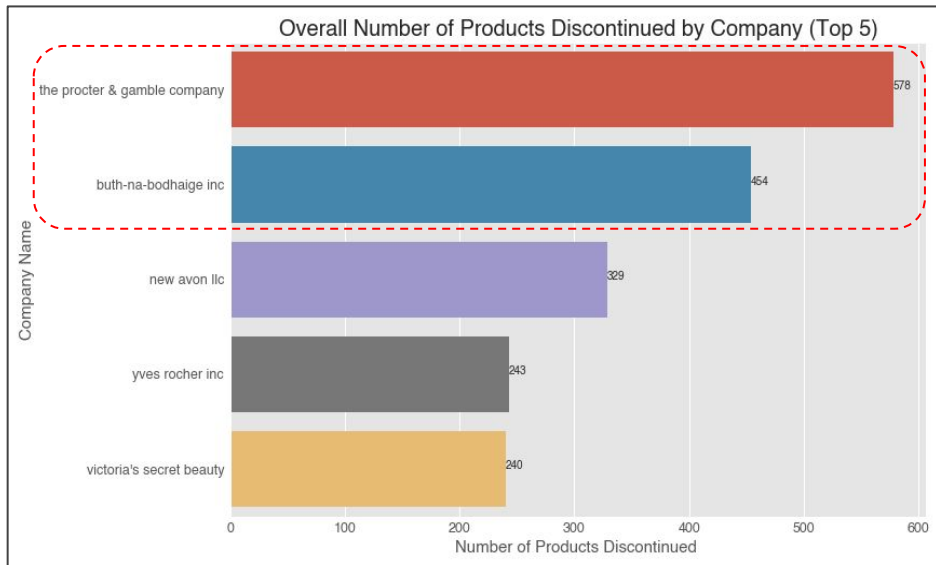
- Two peaks in July 2013 and Oct 2016





Comparing Product Discontinuations

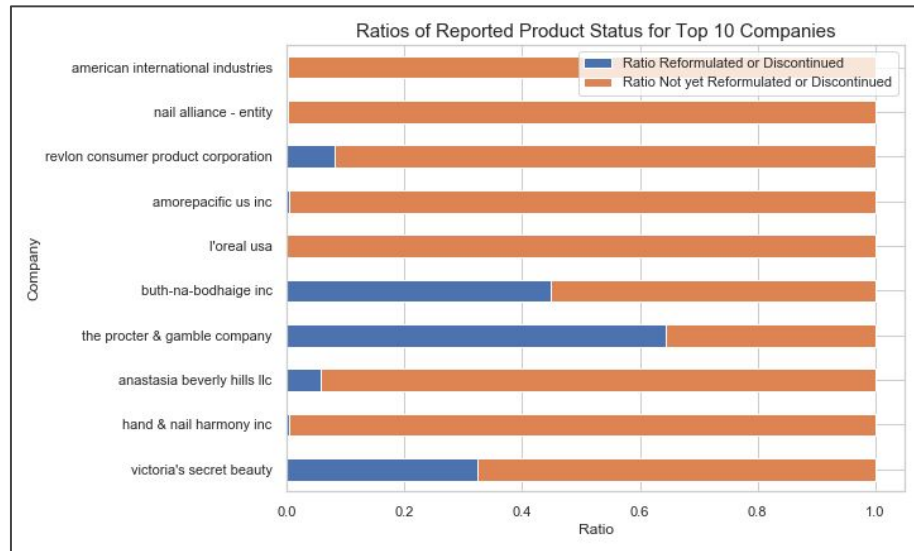
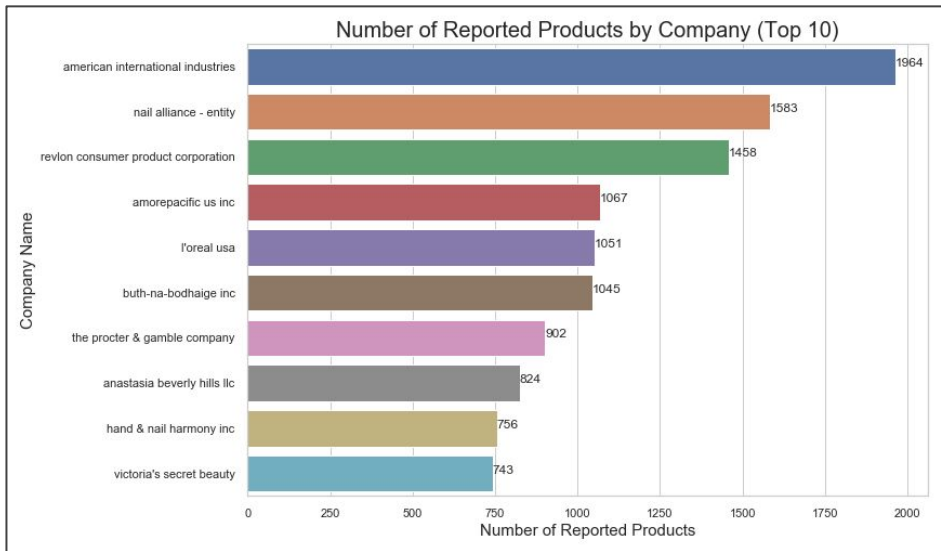
1. **Buth-Na-Bodhaige:** 454/1045 (43%) discontinued
2. **Proctor & Gamble:** 578/902 (64%) discontinued
3. May take over 2-3 years for products to be discontinued





Comparing Product Status Ratios

- Half the top 10 companies have a ratio of reformulated or discontinued products close to 0





Hypothesis Testing



Hypothesis Test #1

Ratio Reformulated or Discontinued

- **Null Hypothesis:**

There is **no** statistically significant difference in the ratio reformulated or discontinued for two companies:

$$P_{\text{Revlon}} = P_{\text{Opi}}$$

- **Alternate Hypothesis:**

There **is** a statistically significant difference in the ratio for two companies:

$$P_{\text{Revlon}} \neq P_{\text{Opi}}$$

- **Threshold = 0.05**
- **P-value = 0.08**
- **Failed to reject null**

Two Sample Approximate Test of Population Proportions

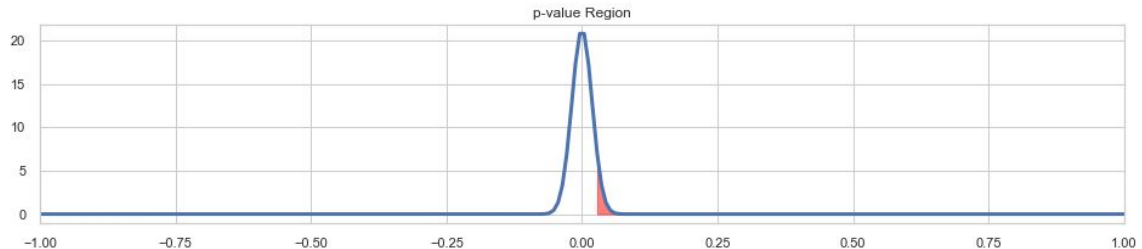
2 Big Nail Product Companies

1. Revlon

- a. # Reformulated/Discontinued = 118
- b. # Samples = 1,458
- c. Ratio = 0.081

2. OPI

- a. # Reformulated/Discontinued = 13
- b. # Samples = 236
- c. Ratio = 0.055



Hypothesis Test #2

Ratio Reformulated or Discontinued

- **Null Hypothesis:**

There is **no** statistically significant difference in the ratio reformulated or discontinued for two companies:

$$P_{\text{Chanel}} = P_{\text{Estee Lauder}}$$

- **Alternate Hypothesis:**

There **is** a statistically significant difference in the ratio for two companies:

$$P_{\text{Chanel}} \neq P_{\text{Estee Lauder}}$$

- **Threshold = 0.05**
- **P-value = 0.02**
- **Reject null hypothesis**

Two Sample Approximate Test of Population Proportions

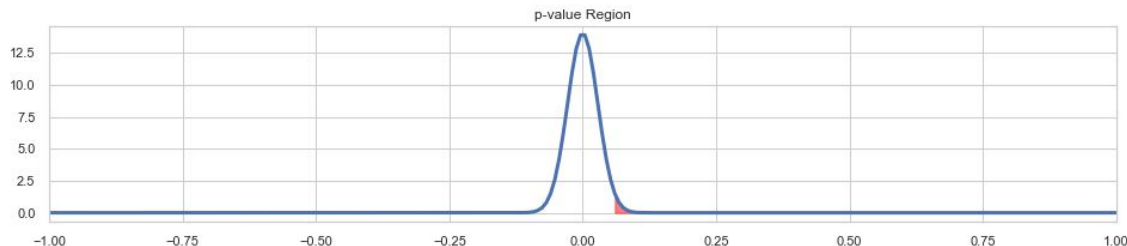
2 Big High-End Product Companies

1. Chanel

- a. # Reformulated/Discontinued = 5
- b. # Samples = 88
- c. Ratio = 0.057

2. Estee Lauder

- a. # Reformulated/Discontinued = 0
- b. # Samples = 68
- c. Ratio = 0





Conclusion

- Majority (85%) of reported products have not yet been reformulated or discontinued
- On average, it took over 2-3 years for products to be discontinued
 - Even for 2 companies that discontinued the most products
- There was **no** statistically significant difference between 2 big nail product companies
- There **was** a statistically significant difference between 2 big high-end companies

Q: Do certain companies handle having harmful cosmetics products much differently?

A: Some may.

- Something to consider: Dataset contained only harmful products
- Future work: Additional domain knowledge/data would help to classify most harmful products vs least harmful



Q&A

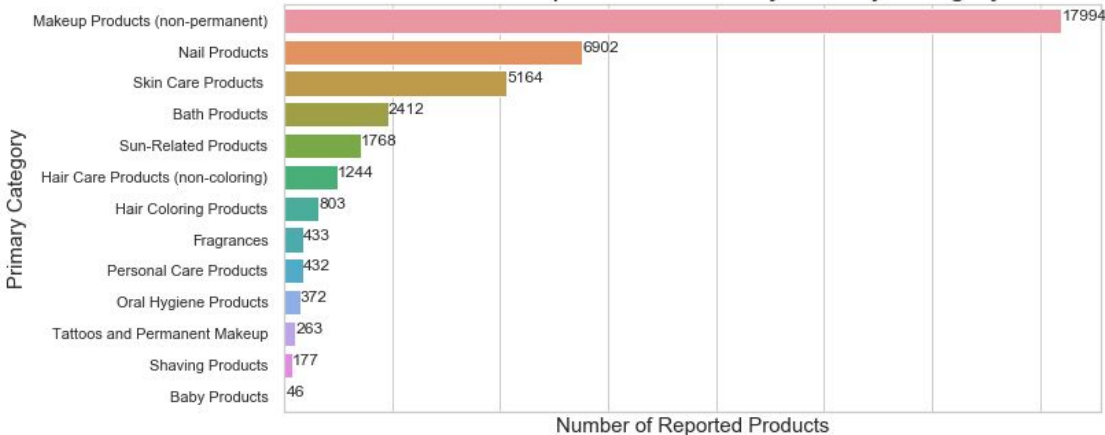


Thank You

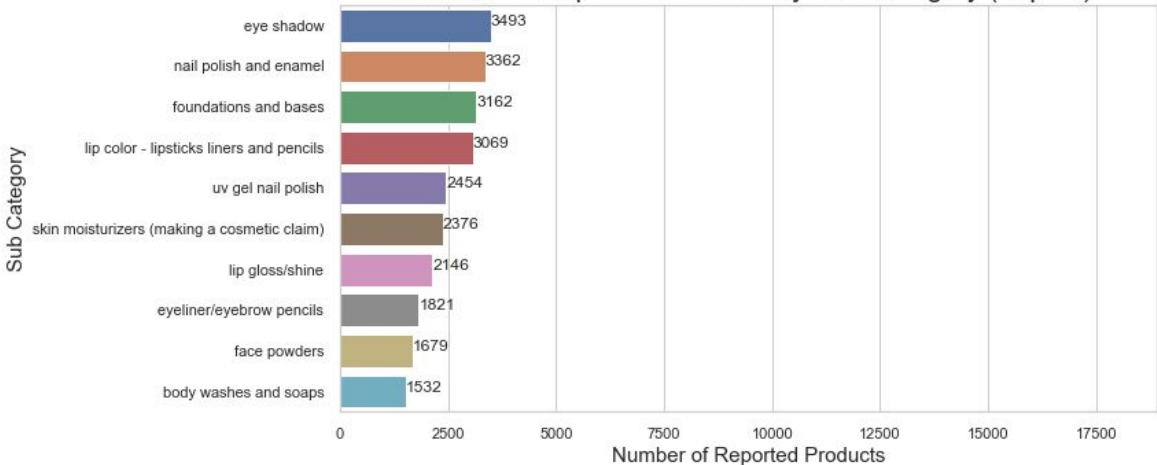
Appendix



Number of Reported Products by Primary Category



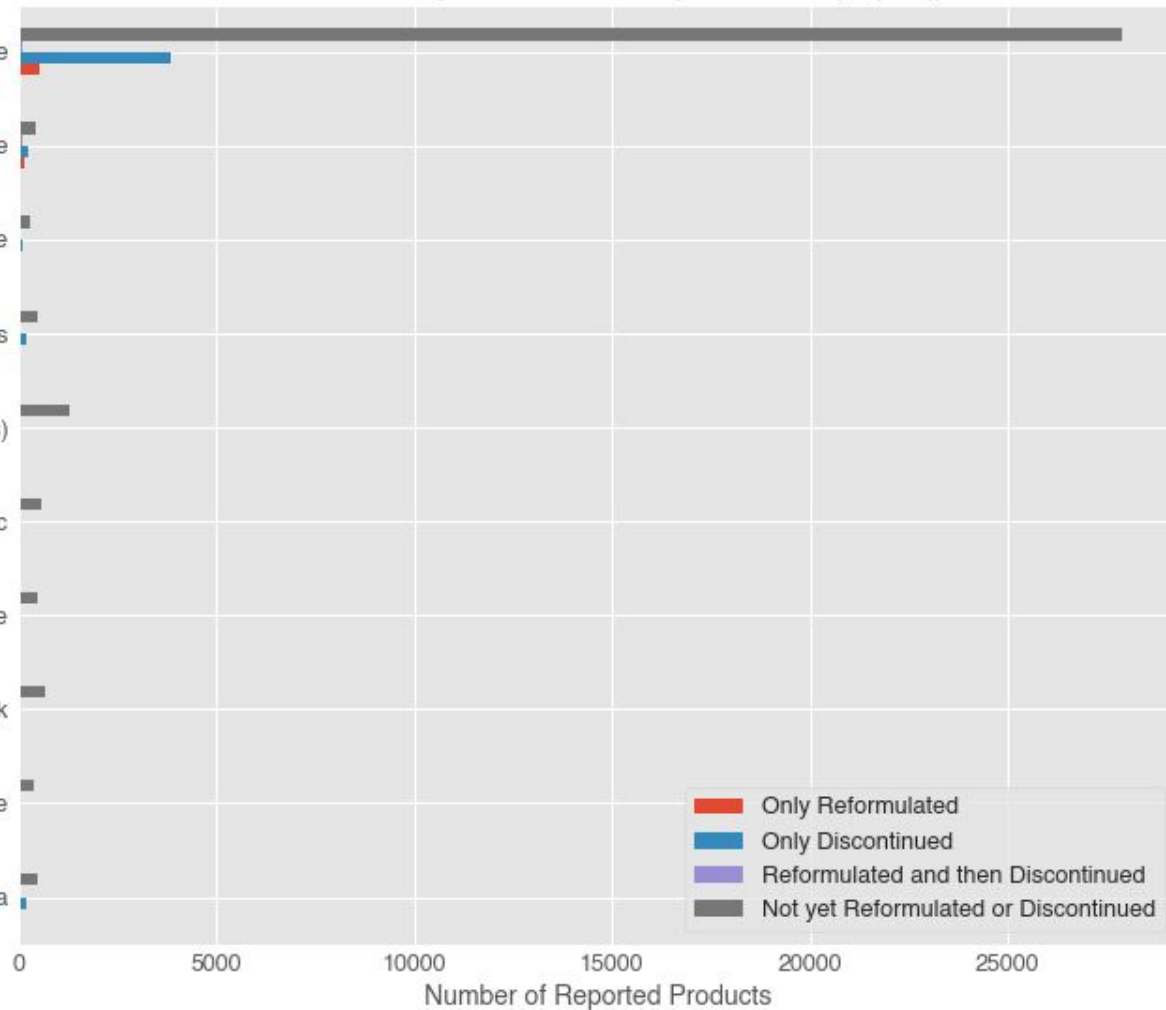
Number of Reported Products by Sub Category (Top 10)



Number of Reported Products by Chemical (Top 10)

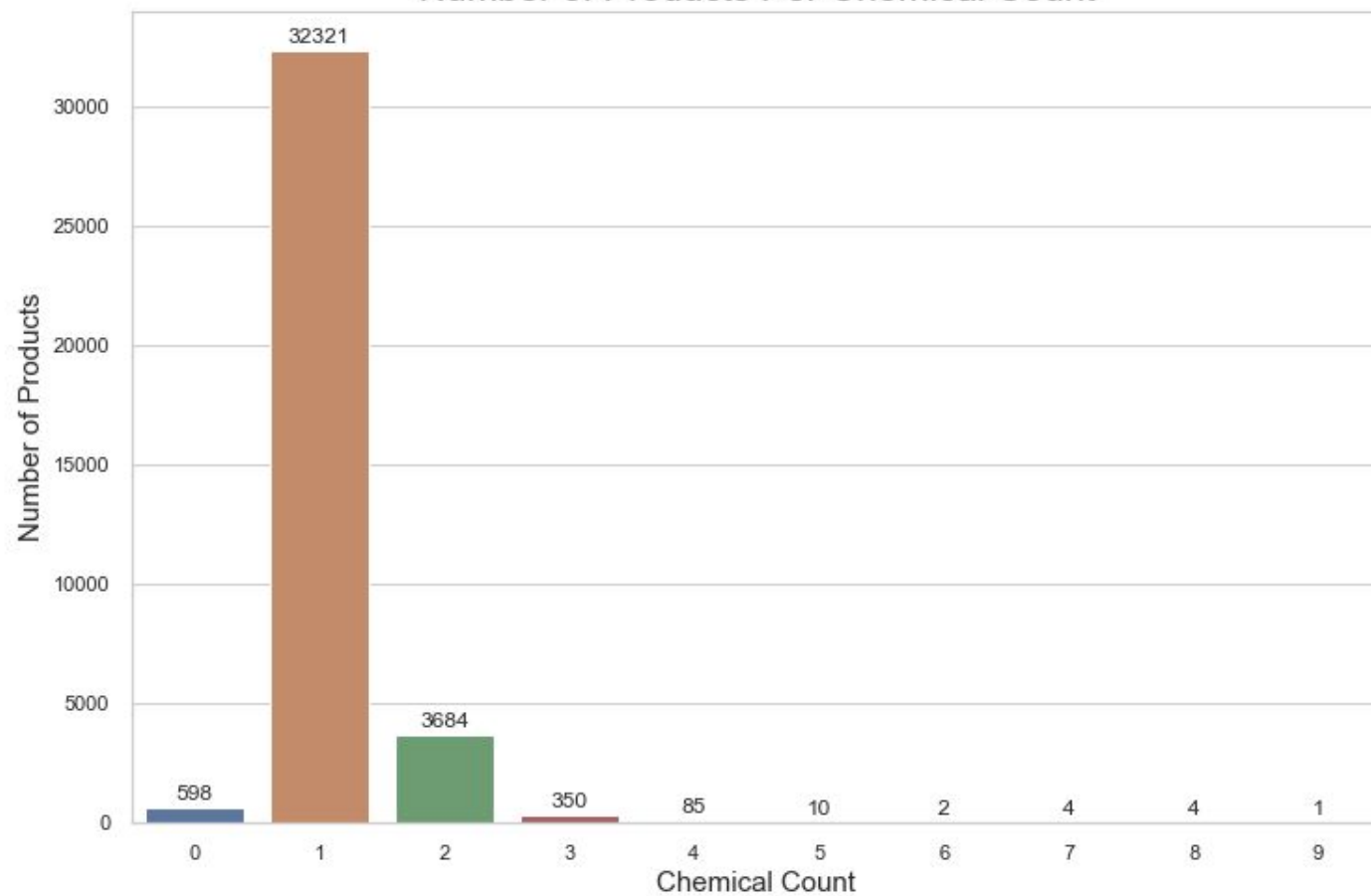
Chemical Name

Titanium dioxide
Cocamide diethanolamine
Butylated hydroxyanisole
Retinol/retinyl esters
Silica, crystalline (airborne particles)
Talc
Retinyl palmitate
Carbon black
Vitamin A palmitate
Mica



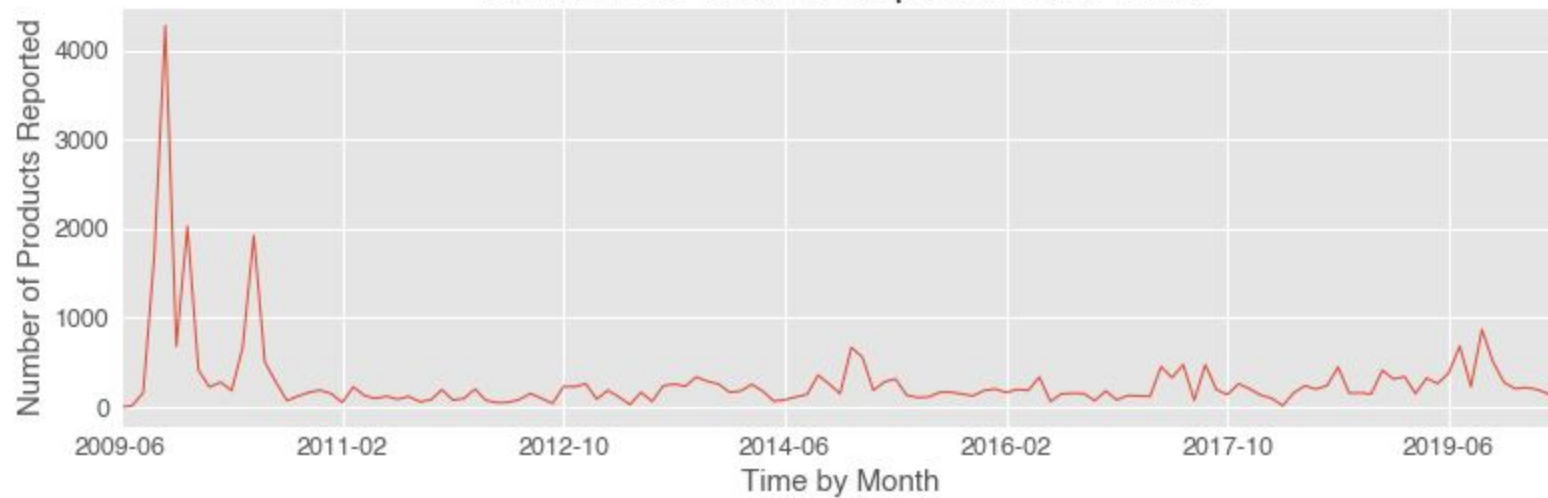
Only Reformulated
Only Discontinued
Reformulated and then Discontinued
Not yet Reformulated or Discontinued

Number of Products Per Chemical Count





Number of Products Reported Over Time



Hypothesis Test:

Time to Product Discontinuation

- Null Hypothesis:

The Proctor & Gamble Company's time to product discontinuation is **less than** Buth-Na-Bodhaige Inc's.

- Alternate Hypothesis:

The Proctor & Gamble Company's time to product discontinuation is **greater than** Buth-Na-Bodhaige Inc's.

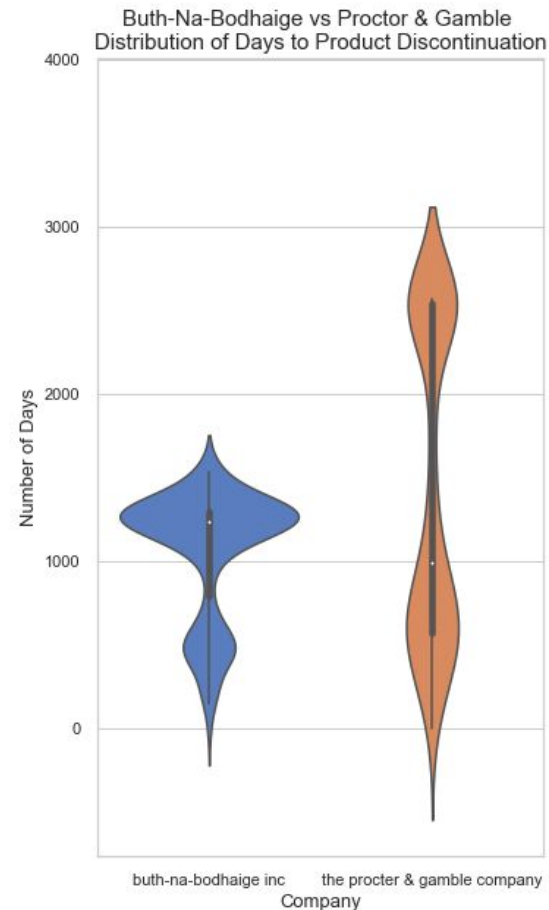
- Threshold = 0.05

- P-value for Proctor & Gamble < Buth-Na-Bodhaige:

- Welch's T-test p-val = $7.4 \times 10^{-14} \sim 0$

- Mann-Whitney U Test p-val = 0.003

- *Reject null hypothesis*



Question:
Do certain companies
handle having harmful
cosmetic products
differently?