Optimizing Cosmetic Formulations Using

Product Reformulation Trends

Olivia Folly-Gah

Project Overview

The cosmetics industry is constantly evolving, with companies reformulating products to address safety concerns, meet regulations, and align with consumer demand for clean beauty. This project explores trends in these reformulations to improve cosmetic formulation strategies.

Table of contents

01

Introduction

Background and Goals

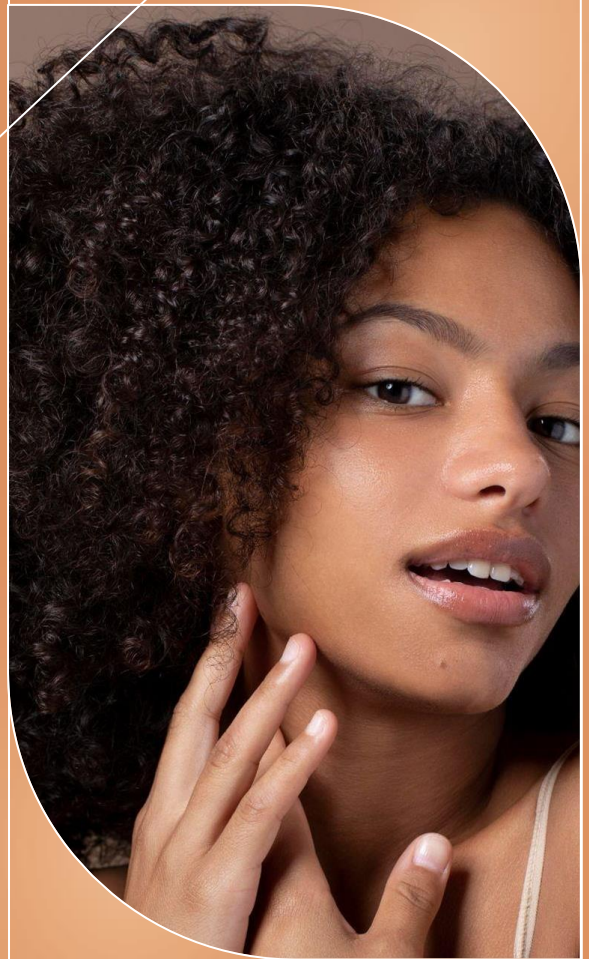
Why Focus on Cosmetics?

Overview of the Industry:

- The cosmetics market is valued at over $500 billion globally and impacts millions of consumers daily.

Importance of Safety in Cosmetics:

- Safety is regulated by laws like the California Safe Cosmetics Act, requiring transparency for hazardous ingredients.
- Reformulations arise due to regulatory updates, safety concerns, and consumer demands for cleaner products.



Objective

Enhancing Cosmetic Formulation Strategies

Analyze cosmetic reformulation trends using stats and machine learning to uncover safety, regulatory, and market-driven patterns.

Relevance to Data Science

Tracks ingredient reformulations with stats and machine learning to predict trends and aid safer, innovative cosmetics.

02

Data Science Approach

Cleaning Process and Challenges

Data Cleaning

Challenges

Missing reformulation dates and inconsistent chemical or company names (e.g., typos, synonyms).

Actions Taken

Standardized chemical names with CAS Numbers.
Mapped CAS Numbers to Prop 65 for hazard scores.
Calculated reformulation frequency from chemical updates.

Example Features

Hazard Score: "Methylparaben: 7, Formaldehyde: 9."
Reformulation: "Yes/No."

Before and After

Column Name	Before Cleaning (Original)	After Cleaning (Cleaned)
ChemicalName	"Methylparaben, Methyl 4-hydroxybenzoate"	Methylparaben
	“Sodium Chloride (Salt)”	Sodium Chloride
CompanyName	“Glow Cosmetics, LLC"	Glow Cosmetics

Before Cleaning

Redundant chemical names with full descriptions

Inconsistent naming conventions (e.g., company suffixes such as "Inc." vs "Ltd." vs "LLC").

After Cleaning

Removed suffixes, extra punctuations, and spaces.

Feature Engineering

01

ChemicalName

Jupiter is the biggest planet of them all

02

HazardType

Categorize chemicals based on regulations.

03

Chemical Age

Calculate the age of each chemical from its creation date to the current date

04

ReformulationCount

Track how many times a product has been reformulated

Exploratory Data Analysis (EDA)....

Makeup

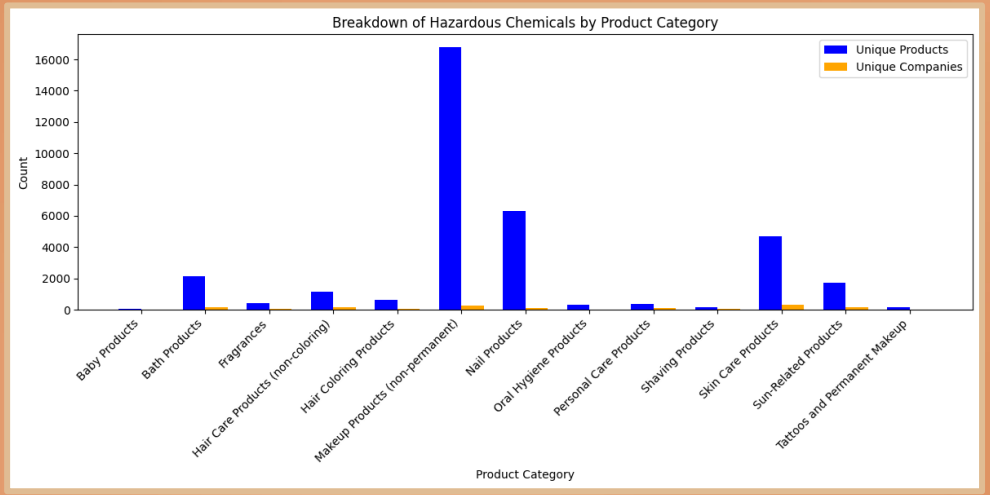
Shows highest discontinuation trend, likely due to bans (e.g., pigments, talc) and changing consumer preferences.

Timeline of Discontinuations

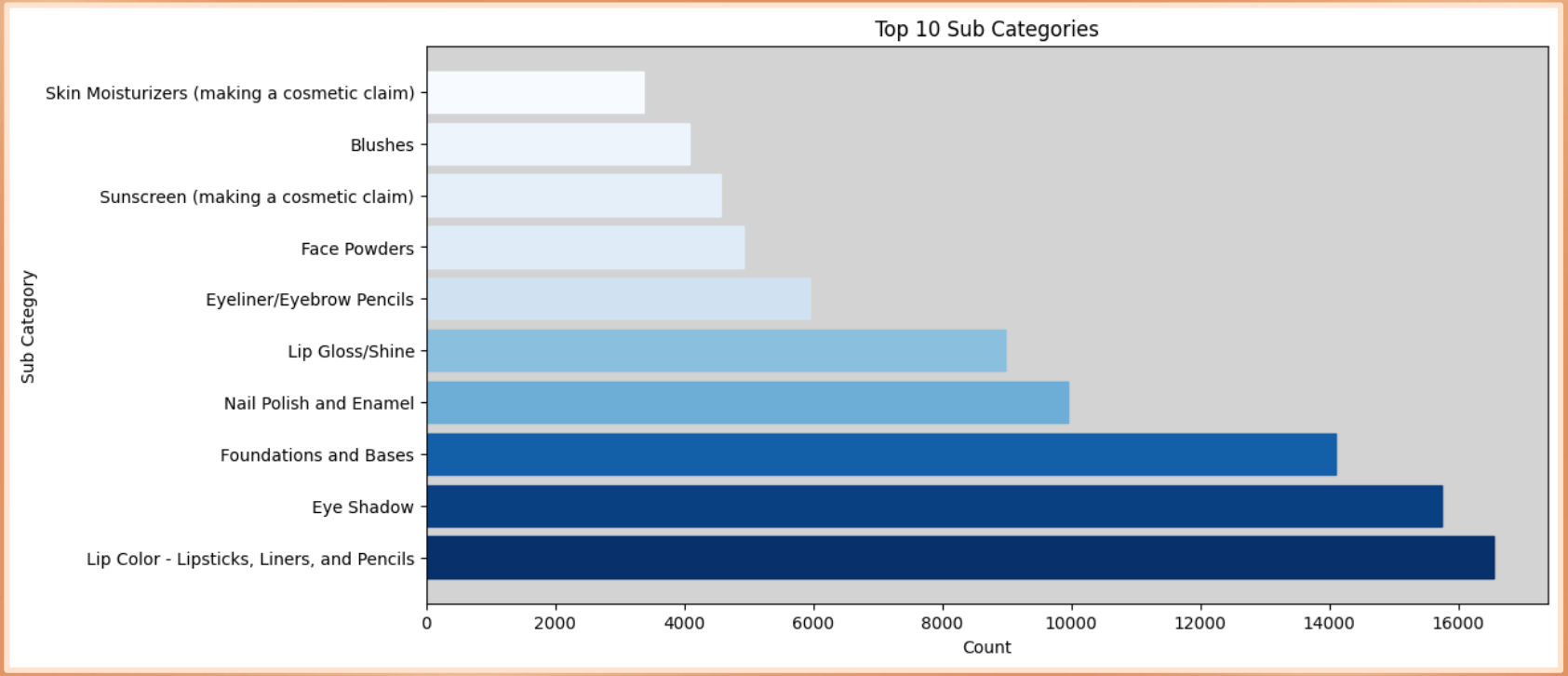
Discontinuation peaks align with regulatory changes and consumer demand for transparency.

Baby Products

Lowest discontinuation rate, reflecting stricter safety standards in initial formulation.



Ranking of the top 10 cosmetic subcategories by the presence of hazardous chemicals.



Adapting to Change: Challenges in Product Discontinuations

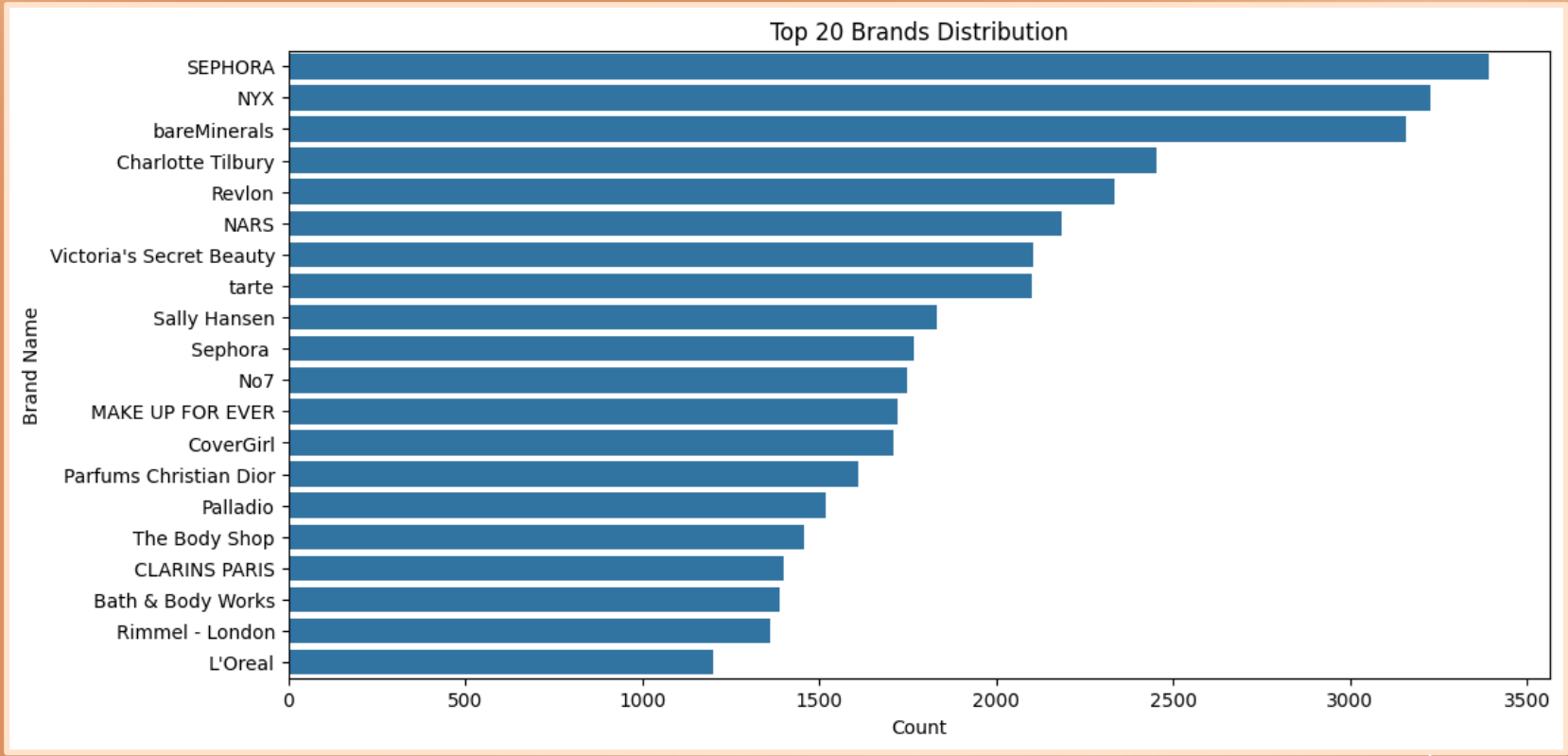
Makeup Products

- Driver: Regulatory scrutiny over ingredients like talc and synthetic pigments.
- Impact: Makeup leads in discontinuations, with smaller brands struggling to adapt due to limited R&D resources.

Baby Products

- Driver: Stringent safety standards in initial formulations.
- Impact: Lowest discontinuation rate, highlighting their proactive compliance with safety expectations.

Top 20 Brands in Cosmetic Dataset



Why Popular Brands Lead in Discontinuations

Key Findings:

1. Large product portfolios increase discontinuation likelihood.
2. Proactive response to regulations and consumer scrutiny.
3. Strategic discontinuations to innovate and rebrand

Conclusion:

- Leading brands use discontinuation as a tool for adaptation, ensuring they remain competitive and aligned with market demands.

Correlations Between Reformulations and Regulations

Key Questions

- Are products with banned chemicals more likely to be reformulated?
- What is the lag time between a regulation update and reformulation?

Correlation:

- Products containing chemicals classified as confirmed carcinogens (e.g., coal tar distillates, titanium dioxide) show a moderate positive correlation with reformulation likelihood (r = 0.70).

Lag Time:

- The lag time between regulation updates and product reformulations also shows a moderate positive correlation (r = 0.70). This indicates that reformulations typically occur after a regulatory update, with a delay of around 6-12 months.

Table of Correlation

Metric	Correlation (r)
Hazard Score ↔ Reformulation Likelihood	0.70
Lag Time ↔ Reformulation Likelihood	0.70

~ **Delayed Responses:** Reformulations tend to follow regulatory changes with a notable delay, suggesting that companies take time to adjust products after regulations are updated.

Applying Naive Bayes Classification to Cosmetic Formulations

Objective

Classify products as Safe vs Hazardous or Reformulated vs Not Reformulated based on ingredient data.

Dataset

Features: Ingredients (binary presence/absence) and hazard scores.
Target: Reformulation status or safety category (Safe, Moderate, Hazardous).

Naive Bayes Model

Predicts reformulation or hazard likelihood based on ingredient presence.
Hazardous chemicals increase reformulation chances.

Advantages

Efficient and scalable for large datasets.
Probabilistic outputs offer insights into the likelihood of reformulation or safety.

Challenges Encountered and Problem-Solving

Challenge 1: Shifting Objective

- Original Goal: Develop safer cosmetic formulas for marginalized communities with curly/coily hair and melanated skin.
- Issue: The dataset lacked specific data on these communities and their unique needs.
- Solution: Shifted focus to analyzing reformulation trends and ingredient safety, adding numerical columns for analysis.

Challenge 2: Ingredient Data Complexity

- Issue: Inconsistent chemical names and CAS numbers.
- Solution: Standardized the data for clearer analysis.

Challenge 3: Lack of Regulatory Data

- Issue: Missing regulatory information to understand ingredient changes.
- Solution: Supplemented with external regulatory databases.

03

Domain Insights

Connecting the analysis with the cosmetic industry

Insights from Cosmetic Chemistry Research

Regulatory Influence on Reformulations

Global regulations, such as EU bans on harmful chemicals (e.g., parabens, phthalates), force companies to reformulate products to meet safety standards.

Consumer Demand for Safety

The "clean beauty" movement has driven consumer demand for transparency, encouraging brands to remove harmful chemicals and prioritize safer formulations.

Industry Challenges in Reformulation

Cosmetic companies face challenges balancing product safety with performance and navigating the complexities of substituting ingredients without compromising efficacy.

04

Future Steps

Plan for Spring Semester

Implementation Plan

Integrate Full Dataset:

- Incorporate the dataset from my industry partner, which now includes more numerical columns (e.g., pH levels, usage rates), allowing me to return to my original goal of developing safer cosmetic formulas for marginalized communities.

Interactive Tableau Visuals:

- Before diving into machine learning, I'll focus on creating an interactive Tableau dashboard to visualize key findings from this semester, highlighting reformulation trends and ingredient safety.

Machine Learning Implementation:

- After visualization, I'll apply clustering and classification techniques to explore product reformulation patterns and identify factors driving ingredient changes.

Work Cited

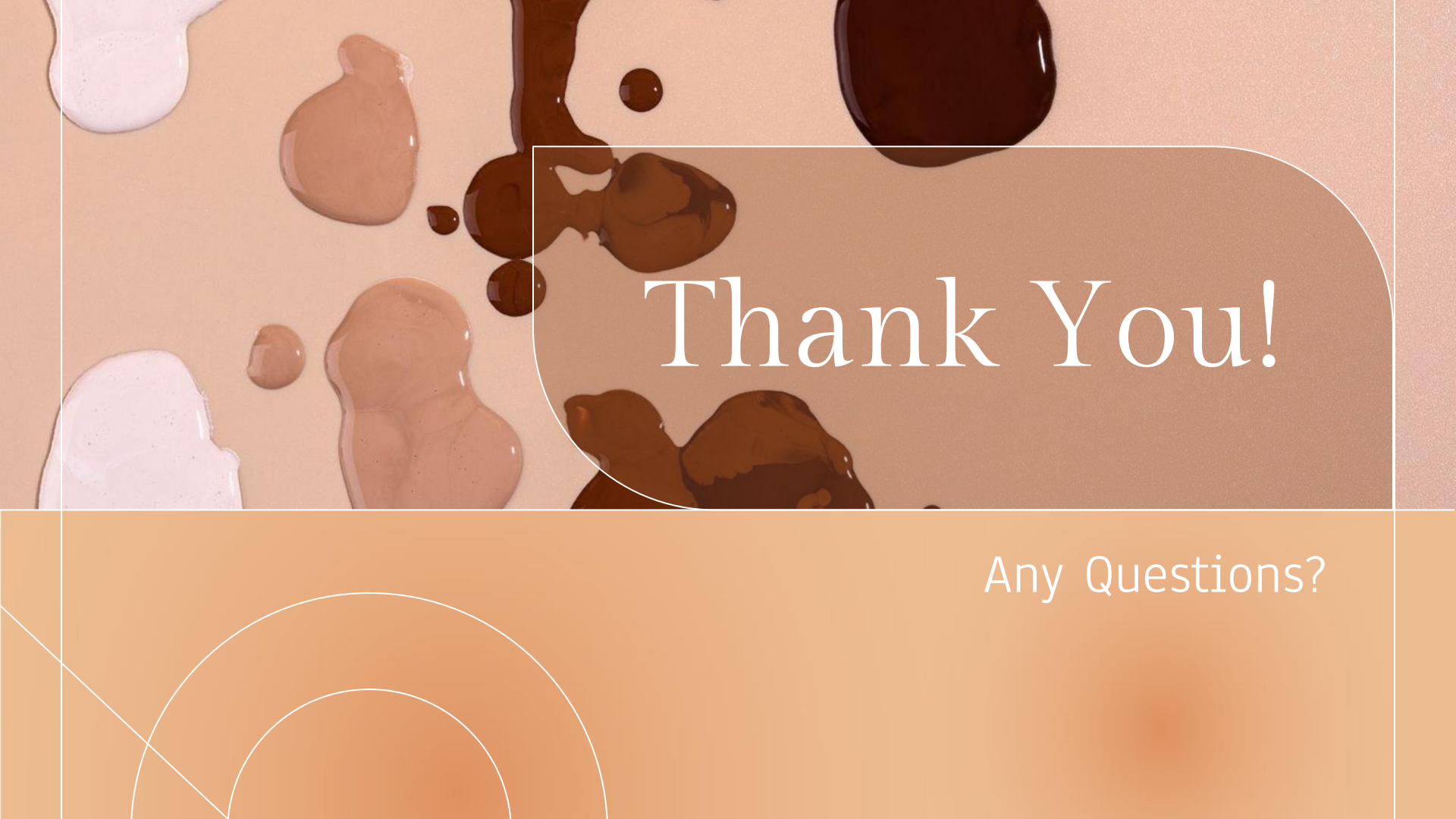
Major for College: Cosmetology | Google Slides and PowerPoint. slidesgo.com/theme/community-family-personal-services-major-for-college-cosmetology#search-cosmetic&position-4&results-206&rs=search&rs=search.

Rosenstein, Olivia Fleming &. Jenna. “The Ultimate Guide to Clean Beauty.” *Harper’s BAZAAR*, 22 Apr. 2020, www.harpersbazaar.com/beauty/skin-care/a28352553/clean-beauty.

“Scientific Opinion on the Hazard Assessment of Endocrine Disruptors: Scientific Criteria for Identification of Endocrine Disruptors and Appropriateness of Existing Test Methods for Assessing Effects Mediated by These Substances on Human Health and the Environment.” *EFSA Journal*, vol. 11, no. 3, Mar. 2013, doi:10.2903/j.efsa.2013.3132.

“---.” *EFSA Journal*, vol. 11, no. 3, Mar. 2013, doi:10.2903/j.efsa.2013.3132.

Xu, Shaohua, et al. “Hair Chemicals May Increase Breast Cancer Risk: A Meta-analysis of 210319 Subjects From 14 Studies.” *PLoS ONE*, vol. 16, no. 2, Feb. 2021, p. e0243792, doi:10.1371/journal.pone.0243792.


Thank You!
Any Questions?