Analysis of Chemical Components

badhusha sathar

2025-03-22

Tableau Visualizations:

CLICK (https://public.tableau.com/app/profile/badhusha.sathar/viz/COMPONENTANALYSIS)

ASK

Overview

This dataset provides structured information on cosmetic ingredients, product rankings, price comparisons, and effects on different skin types. The primary goal of the analysis is to enhance consumer understanding of product composition, enabling informed skincare choices.

The dataset serves as a valuable resource for understanding chemical composition, skincare preferences, and pricing strategies. Future expansions could incorporate ingredient safety evaluations, user reviews, and sustainability indicators for a more comprehensive analysis.

PREPARE

- The dataset used for the case study is from kaggle.click here (https://www.kaggle.com/datasets/kingabzpro/cosmetics-datasets).

- The cosmetic dataset appears to be in a wide format, rather than long format. In the dataset, each row represents a single cosmetic product, with multiple attributes (such as brand, price, rank, ingredients, and skin type suitability) stored in separate columns

- The dataset is licensed under GPL 2, meaning it is open-source and can be used for research and analysis.

- The dataset does not contain personally identifiable information (PII), ensuring user privacy.

- It focuses on cosmetic product attributes rather than individual consumer data.

- The dataset is hosted on Kaggle, a secure platform for data sharing.

- The dataset is publicly available and can be accessed by anyone interested in cosmetic ingredient analysis.

PROCESS

- Tools used - bigquery,tableau

```
      #check first five rows of the table cosmetics
SELECT *
FROM `chemicalcomponent.component_analysis.cosmetics`;
LIMIT 5;

# CHECK RANDOM ROWS OF THE TABLE

SELECT *
FROM `chemicalcomponent.component_analysis.cosmetics`
ORDER BY RAND()
LIMIT 5;

# CHECK FOR DUPLICATES
WITH duplicates AS (
  SELECT *,
         ROW_NUMBER() OVER (PARTITION BY Label, Brand, Name, Price, Rank, Ingredients, Combin
ation, Dry, Normal, Oily, Sensitive ORDER BY Label) AS row_num
  FROM `chemicalcomponent.component_analysis.cosmetics`
)
SELECT *
FROM duplicates
WHERE row_num > 1;
## NO DUPLICATES FOUND

# CHECK TOTAL NUMBER OF ROWS
SELECT COUNT(*) AS total_rows
FROM `chemicalcomponent.component_analysis.cosmetics`;
##RESULT:1472

# check number of columns
SELECT COUNT(*) AS total_columns
FROM `chemicalcomponent.component_analysis.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'cosmetics';

##RESULT:11

# CHECK FOR NULL VALUES
SELECT *
FROM `component_analysis.cosmetics`
WHERE
  Label IS NULL OR
  Brand IS NULL OR
  Name IS NULL OR
  Price IS NULL OR
  Rank IS NULL OR
  Ingredients IS NULL OR
  Combination IS NULL OR
  Dry IS NULL OR
  Normal IS NULL OR
  Oily IS NULL OR
  Sensitive IS NULL;

  # RESULT: NO NULL VALUES FOUND
```

ANALYSIS

TOTAL BRAND

```
--total brands--

SELECT
COUNT(DISTINCT Brand) AS TotalBrands
FROM `chemicalcomponent.component_analysis.cosmetics`;
```

RESULT [{ "TOTALBRAND": "116" }]

TOTAL LABELS

```
SELECT
COUNT(DISTINCT Label) AS TotalLabel
FROM `chemicalcomponent.component_analysis.cosmetics`;
```

RESULT [{ "TotalLabel": "6" }]

THE LABEL WITH HIGHEST PRICE

```
SELECT Label, MAX(Price) AS HighestPrice
 FROM `chemicalcomponent.component_analysis.cosmetics`
 GROUP BY Label
 ORDER BY HighestPrice DESC LIMIT 1;
```

RESULT: [{ "Label": "Treatment", "HighestPrice": "370" }] ### BRAND WITH HIGHEST PRICE

```
SELECT Brand, MAX(Price) AS HighestPrice
 FROM `chemicalcomponent.component_analysis.cosmetics`
 GROUP BY Brand
 ORDER BY HighestPrice DESC LIMIT 1;
```

RESULT: [{ "Brand": "LA MER", "HighestPrice": "370" }]

```
SELECT Brand

 FROM `component_analysis.cosmetics`
 ORDER BY Rank
 DESC LIMIT 1;
```

RESULT: [{ "Brand": "LA MER" }]

THE TOP 5 BRAND with HIGHEST RATING

```
SELECT Brand, Rank
FROM `component_analysis.cosmetics`
ORDER BY Rank
DESC LIMIT 5;
```

RESULT:

[{ "Brand": "NURSE JAMIE", "Rank": "5.0" }, { "Brand": "LANEIGE", "Rank": "5.0" }, { "Brand": "DIOR", "Rank": "5.0" }, { "Brand": "CLINIQUE", "Rank": "5.0" }, { "Brand": "LA MER", "Rank": "5.0" }]

BRAND NAMES AND LABEL NAMES

```
# DISPLAY DISTINCT BRAND

SELECT DISTINCT Brand
 FROM `chemicalcomponent.component_analysis.cosmetics`;

#DISPLAY DISTINCT LABEL

SELECT DISTINCT Label
 FROM `chemicalcomponent.component_analysis.cosmetics`;
```

NUMBER OF BRAND SUITABLE FOR ALL SKIN TYPE

```
SELECT COUNT(DISTINCT Brand) AS all_skin_types_count
FROM  `chemicalcomponent.component_analysis.cosmetics`
WHERE Combination = 1
  AND Dry = 1
  AND Normal = 1
  AND Oily = 1
  AND Sensitive = 1;
```

RESULT [{ "all_skin_types_count": "91" }]

INGREDIENT EXTRACTION

```
# LIST UNIQUE INGREDIENTS COMBO
SELECT DISTINCT Ingredients
FROM `chemicalcomponent.component_analysis.cosmetics`;



#Standardize Ingredient Names

SELECT TRIM(LOWER(Ingredients)) AS standardized_ingredients
FROM `chemicalcomponent.component_analysis.cosmetics`;



#Split Ingredients into Individual Components

SELECT ingredient
FROM (
  SELECT SPLIT(Ingredients, ', ') AS ingredient_array
  FROM `chemicalcomponent.component_analysis.cosmetics`
), UNNEST(ingredient_array) AS ingredient;



#REMOVE UNWANTED CHARACTERS

SELECT REGEXP_REPLACE(Ingredients, '[^a-zA-Z, ]', '') AS cleaned_ingredients
FROM `chemicalcomponent.component_analysis.cosmetics`;

#Aggregate Standardized Ingredients

SELECT  ingredient,Price,Brand
FROM (
  SELECT SPLIT(Ingredients, ', ') AS ingredient_array,Price,Brand
  FROM `chemicalcomponent.component_analysis.cosmetics`
), UNNEST(ingredient_array) AS ingredient;
```
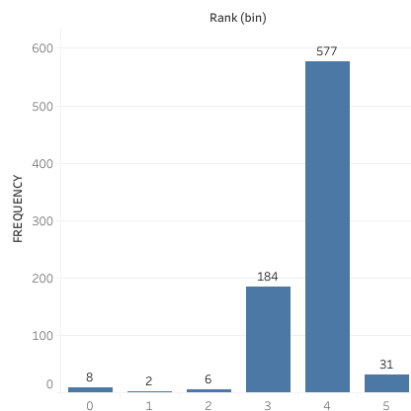
Visualization of Data.

In this phase, we will be visualizing the data analyzed and tables created using Tableau Public.

For the interactive version,Click Here to View Tableau Dashboard (https://public.tableau.com/views/COMPONENTANALYSIS/Dashboard1?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link)
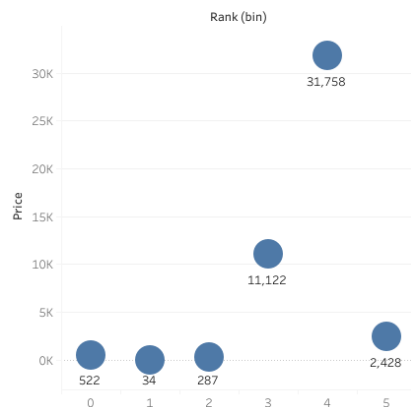
DISTRIBUTION OF PRODUCT RATING

Rank (bin)

KEY FINDINGS: - The histogram shows the frequency of different product ratings.

- Most products seem to cluster around mid-range ratings, with fewer products at extreme high or low ratings.

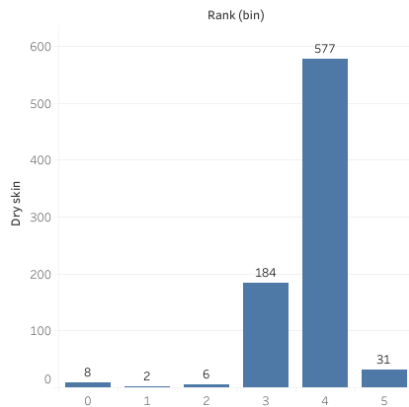## Product Price vs. Rating Distribution

Rank (bin)

key findings: - Higher-priced products tend to have more varied ratings, suggesting price alone doesn't guarantee better reviews.
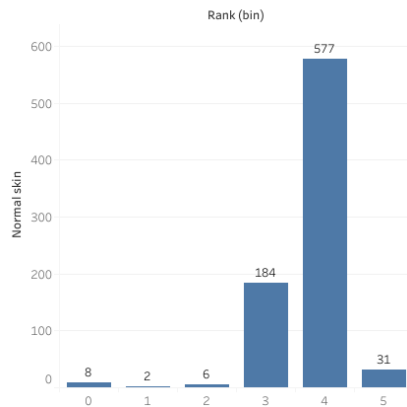
- Some lower-priced products still achieve high ratings, indicating affordability doesn't always mean lower quality.
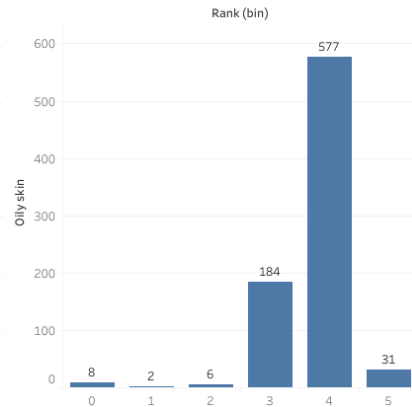
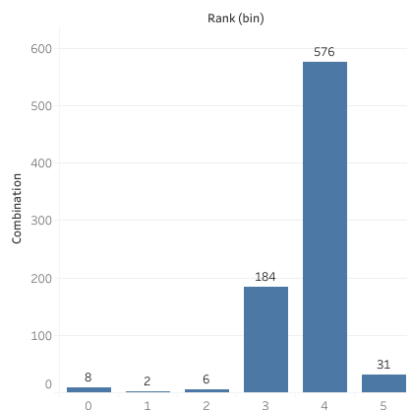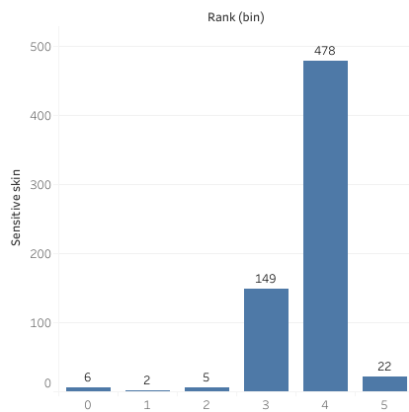# SKIN TYPES AND PRODUCT RATING.

DRY SKIN PRODUCT V/S RATING

Rank (bin)

NORMAL SKIN V/S PRODUCT RATING

Rank (bin)

OILY SKIN V/S PRODUCT RATING

Rank (bin)

SENSITIVE SKIN V/S PRODUCT RATING

Rank (bin)

## combination v/s distribution of rating
(combination=both dry and oily)

Rank (bin)

kEY FINDINGS:

- Higher-ranked products tend to have consistent and balanced ingredient compositions.

- Some lower-ranked products might lack key moisturizing components.

- Some lower-ranked products might lack key hydrating or oil-controlling components.

- Lower-rated products might indicate ingredient gaps or consumer dissatisfaction.

BRAND V/S PRICE
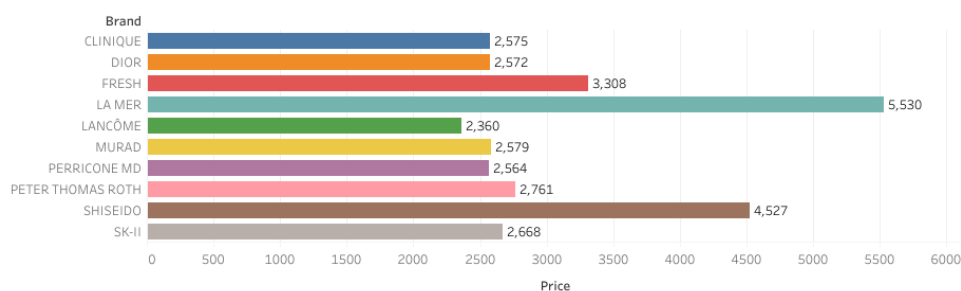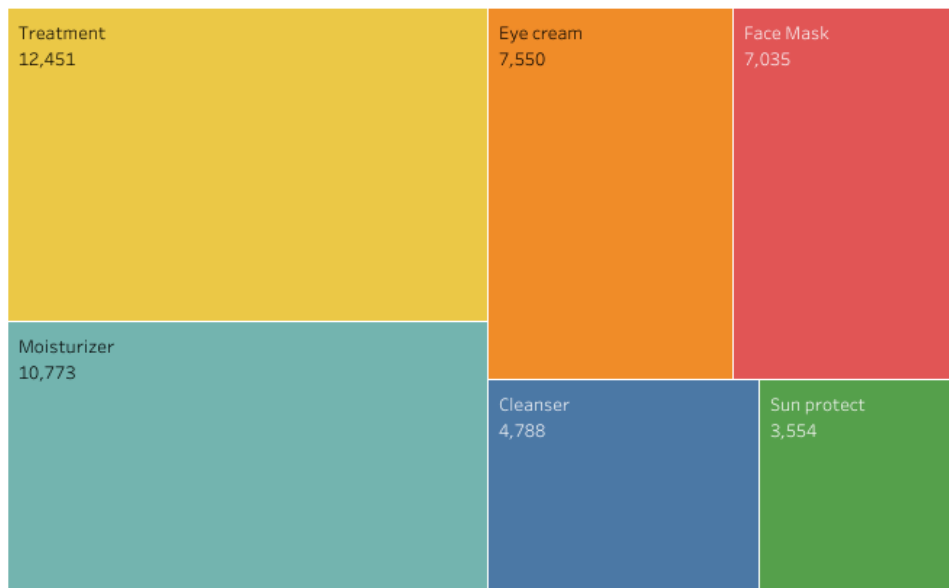

kEY FINDINGS

- La Mer and SK-II appear to have the highest-priced products, while Murad and Peter Thomas Roth have more affordable options.

- Clinique, Dior, and Lancôme fall into the mid-to-high price range, suggesting they cater to both premium and general consumers.

- Shiseido and Perricone MD have a mix of moderate and high-end products, indicating diverse pricing strategies.

- Brands with higher average prices may focus on premium skincare, while lower-priced brands might target mass-market consumers.
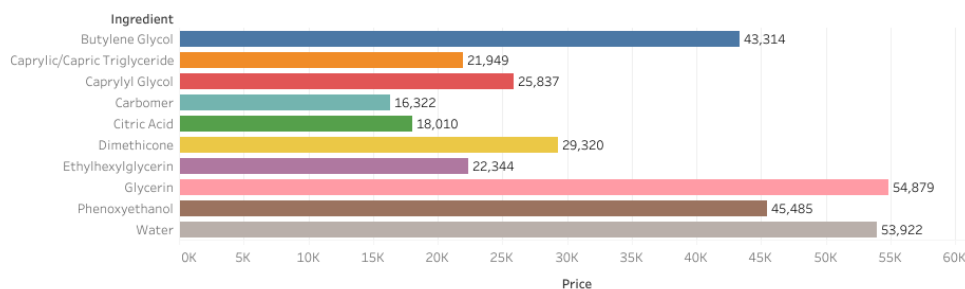

LABEL V/S PRICE

Treatment 12,451	Eye cream 7,550	Face Mask 7,035
Moisturizer 10,773	Cleanser 4,788	Sun protect 3,554

KEY FINDINGS:

- Moisturizers and Treatments appear to have the highest-priced products, while cleansers and sunscreens are generally more affordable.

- Eye creams and face masks tend to be in the mid-to-high price range, suggesting they cater to both premium and general consumers.

- Sun protection products are priced lower, possibly due to mass-market availability.

- Categories with higher average prices may focus on premium skincare, while lower-priced labels might target daily-use consumers.

INGREDIENT V/S PRICE

Ingredient	Price
Butylene Glycol	43,314
Caprylic/Capric Triglyceride	21,949
Caprylyl Glycol	25,837
Carbomer	16,322
Citric Acid	18,010
Dimethicone	29,320
Ethylhexylglycerin	22,344
Glycerin	54,879
Phenoxyethanol	45,485
Water	53,922

KEY FINDINGS :

- Butylene Glycol, Caprylic/Capric Triglyceride, and Dimethicone appear to be among the most frequently used ingredients.

- Water is present in almost all products, but its price impact is minimal.

- Some ingredients like Phenoxyethanol and Ethylhexylglycerin are found in both high-priced and low-priced products, indicating they are widely used across different formulations.

- Luxury skincare products tend to include rare botanical extracts, which may contribute to higher costs.

- Ingredients with complex chemical compositions or rare natural sources tend to be in the higher price range.

- Commonly used preservatives and emulsifiers are found in affordable products, suggesting they are cost-effective.

Ingredient Usage Across Brands
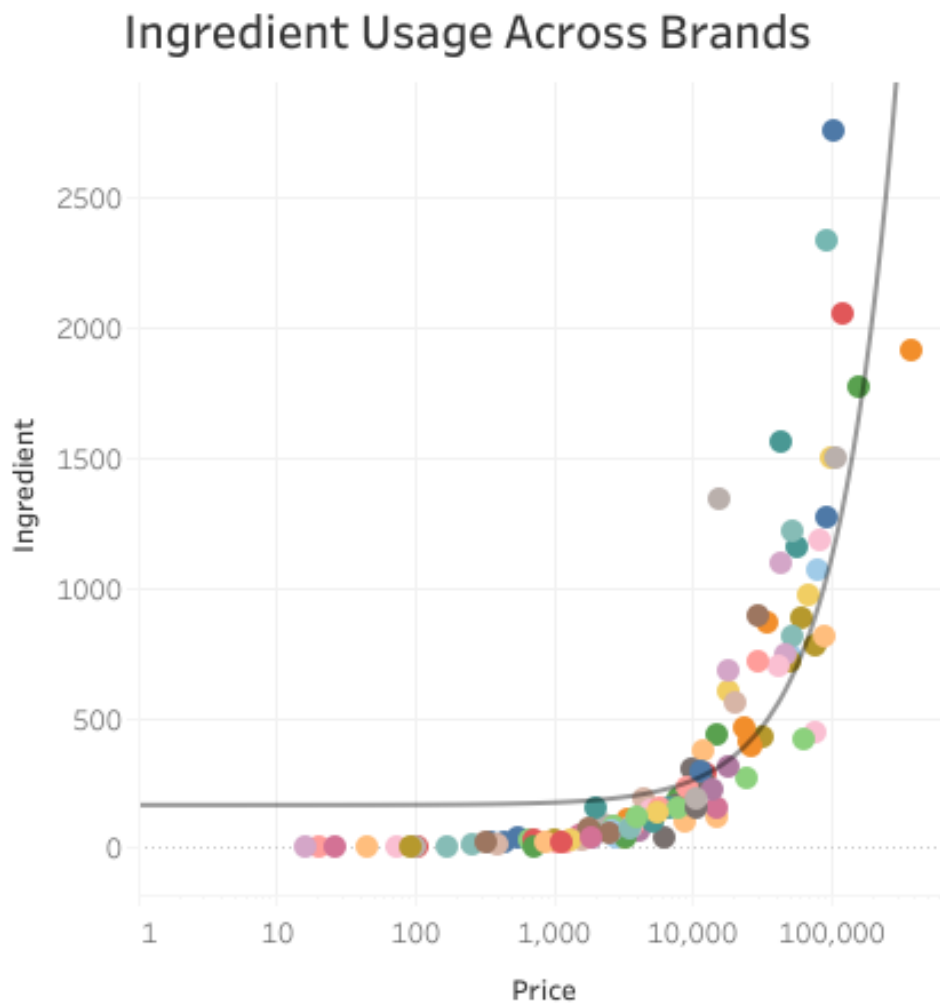
KEY FINDINGS:

- Some brands rely heavily on common preservatives, while others focus on premium botanical extracts.
- Certain ingredients appear more often in high-rated products, suggesting they contribute to better consumer satisfaction.
- Brands using rare or specialized ingredients may position themselves as luxury skincare providers.
- The dataset suggests that higher-priced brands tend to include exclusive ingredients.
- More affordable brands use widely available compounds, ensuring cost-effective formulations.

###Tableau Dashboard

Visualizations built in a dashboard.

DISTRIBUTION OF PRODUCT RATING
Rank (bin)

Product Price vs. Rating Distribution
Rank (bin)

DRY SKIN PRODUCT V/S RATING
Rank (bin)

NORMAL SKIN V/S PRODUCT RATING
Rank (bin)

OILY SKIN V/S PRODUCT RATING
Rank (bin)

INGREDIENT V/S PRICE
Ingredient
Butylene Glycol 43,314
Caprylic/Capric Triglyceride 21,949
Caprylyl Glycol 25,837
Carbomer 16,322
Citric Acid 18,010

SENSITIVE SKIN V/S PRODUCT RATING
Rank (bin)

combination v/s distribution of rating
(combination=both dry and oily)
Rank (bin)

BRAND V/S PRICE
Brand
CLINIQUE 2,575
DIOR 2,572
FRESH 3,308
LA MER 5,530
LANCÔME 2,360

LABEL V/S PRICE
Treatment 12,451
Moisturizer 10,773
Eye cream 7,550
Cleanser 4,788
Face Mask 7,035

Ingredient Usage Across Brand

CONCLUSION

The cosmetic product analysis provides valuable insights into ingredient composition, pricing strategies, and product ratings across different skin types.

Ingredient Impact on Pricing – Luxury brands incorporate premium ingredients, leading to higher costs, while budget-friendly brands optimize cost-effective formulations.onsumer Ratings & Skin Type Preferences – Products tailored to dry skin generally receive higher ratings when containing hydrating agents like Glycerin and Dimethicone. Brand Price Variations – Brands like La Mer, SK-II, and Shiseido dominate the high-end segment, whereas Clinique and Murad maintain more accessible pricing. Ingredient Distribution – Certain ingredients, such as Citric Acid and Phenoxyethanol, are widely present across multiple formulations, impacting skin suitability. Combination & Sensitive Skin Trends – Products labeled for combination skin maintain balanced ingredient concentrations, and minimal irritants improve ratings for sensitive skin users.

Top Recommendations to Marketing Strategists:

1. Leverage Ingredient-Based Targeting

- Highlight key skincare ingredients in marketing campaigns to attract ingredient-conscious consumers.
- Example: Promote Glycerin & Hyaluronic Acid for dry skin solutions.

2.Optimize Pricing Strategies

- Adjust pricing based on brand positioning and consumer demand.
- Example: Luxury brands like La Mer & SK-II should focus on premium packaging & exclusivity, while mid-range brands can emphasize cost-effective alternatives.

3.Enhance Personalization with AI Recommendations

- Implement ingredient-similarity-based product recommendations.
- Example: Suggest similar products for consumers based on skincare concerns (dry, oily, sensitive).

4.Increase Consumer Trust Through Transparency

- Showcase ingredient safety ratings and dermatologist-backed insights.

- Example: Highlight products with minimal irritants for sensitive skin users.

5. Expand Multi-Skin-Type Marketing Approach

- Promote multi-functional products that cater to combination skin users.

- Example: Formulate products that work for both oily and dry skin to widen target audience appeal.