January 2024

SEPHORA Skincare Treatments Analysis

Descriptive and Prescriptive Analysis of SEPHORA Products

Python, Microsoft Excel, Tableau

Introduction:

The Sephora CFO is conducting an evaluation of Sephora's skincare treatment product lineup, aiming to optimize performance and customer satisfaction. This entails a price analysis across various brands and identifying underperforming brands for removal, making space for either already successful brands or newcomers.

Furthermore, the CFO seeks insights into the popularity of specific skincare concerns, such as dryness, acne, or oiliness. Understanding these customer needs will guide in product selection and reveal market gaps and opportunities. These insights will inform strategic decisions when introducing new products or brands, ensuring alignment with customer preferences and market demand.

Objectives:

1. Exploratory Price Analysis
2. Brand Performances
3. Market Gaps and Opportunities with Product Solutions

Data Collection:

Using python, I've written two files; one to scrape the product links and one to scrape information from each product into an excel spreadsheet. These files use modules selenium, pandas, and openpyxl. These files can be found in the Github repository.

The products are scraped from: https://www.sephora.com/shop/facial-treatments ; total of approximately 600 products

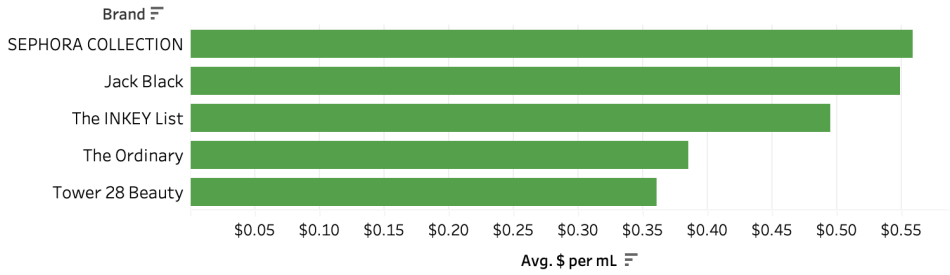
Data collected includes Product Name, Brand, Price, Size in oz. and mL., Rating, Number of Ratings, and Solution.

Data Cleaning and Preprocessing:

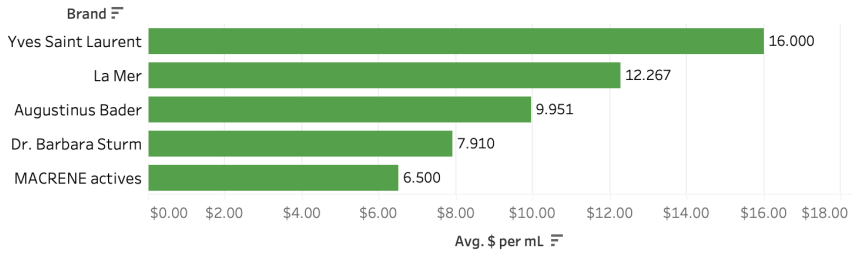
1. Manually Insert data for products not scraped / products that gave errors when scraping
2. Dedupe: check for duplicates
3. Turn dataset into table
4. Handle Missing Values: Delete ‘Products not Carried’
5. Clean ‘Size’ Column to remove unnecessary substring with text splitting.
6. Text split Rating ‘4 stars’ into ‘4’ ‘stars’ using text to columns and delimiter
7. Convert number of ratings into whole numbers, ex. 3.1k into 3100
8. Separate size column into two different columns; one for mL, one for oz PLUS extra column for ‘other’ sizes (ex. 8 Treatments/Pack)
9. Add missing sizes manually through research. Add missing conversions from oz column to mL column.
10. Transposing solutions column into a readable format by using formulas and pivot tables
11. Inconsistent data: Cleanup text for uniformity using ‘Find and Replace’

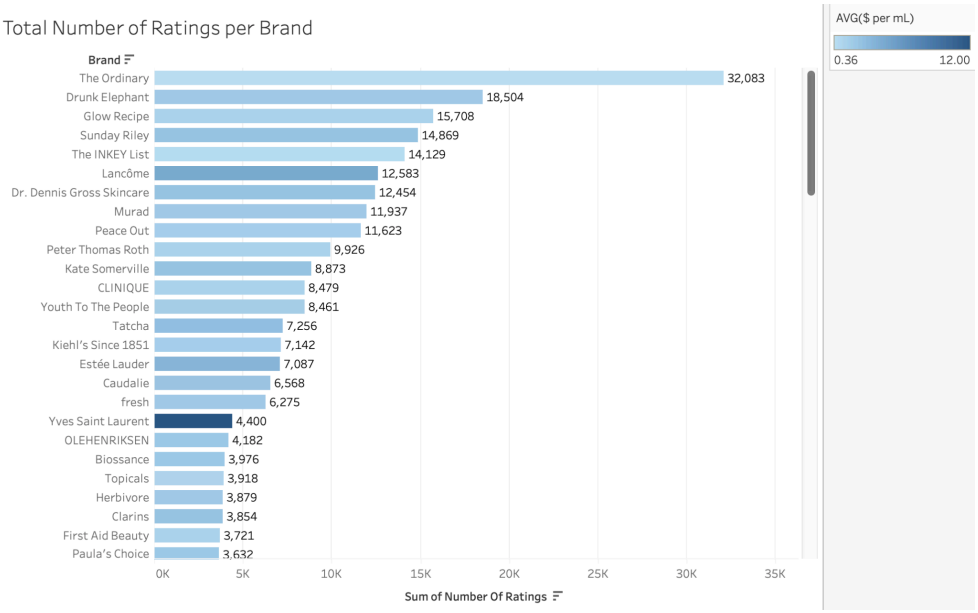
Exploratory Data Analysis:

Brands with lowest prices

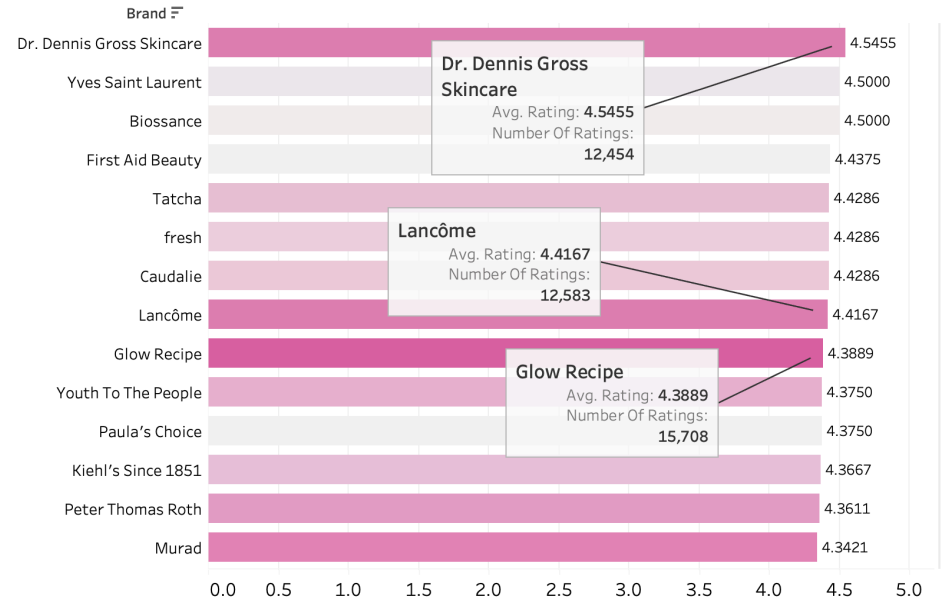


Brands with highest prices

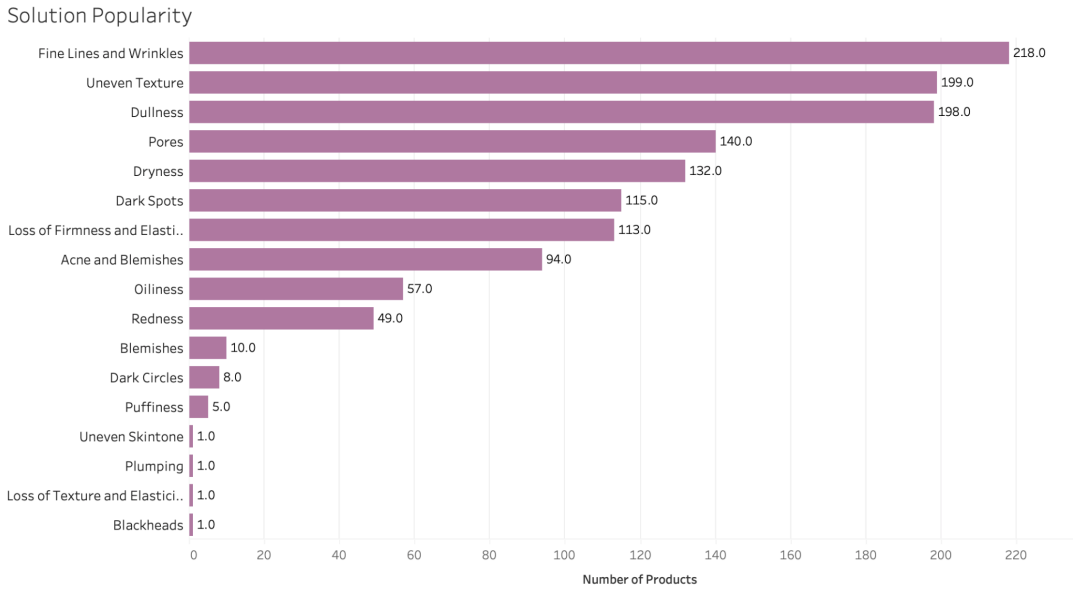




Best performing brands, deeper color means higher number of ratings total



Solution Popularity



_____

Insights: Price ranges from an average of $0.36/mL to $16.00/mL for brands. Cheaper brands doesn't necessarily mean higher popularity (measured in total number of ratings).

Dr. Dennis Grossman Skincare, Lancome, and Glow Recipe are very well performing brands; they have high ratings along with a high number of ratings, deeming them as very successful brands on Sephora shelves.

Low number of products for specific solutions can indicate a market opportunity, but customer preferences for products must be analyzed next.

Discussion:

Strength and Limitations:

Data was gathered by web scraping of Sephora's website (www. sephora. com). The analysis conducted relies solely on this data and lacks access to current sales and other pertinent data, which may affect the accuracy of the findings.

Future Research:

The next steps involve assessing underperforming brands to decide what stays and what goes, making space for new brands and products. Additionally, conducting a customer profile analysis will guide decisions on introducing new products tailored to our market.