PRODUCT

SI 618 Project -- Sephora Data Analysis

Group 23
Ruizhe Wang, Zhiyi Ji
12/06/2024

INTRODUCTION

Analysis of Sephora Products and Skincare Reviews Dataset

Objectives:

- Identify factors driving high product ratings and satisfaction.
- Explore time trends in consumer satisfaction and category preferences.
- Link specific ingredients to customer satisfaction or dissatisfaction.

Approach:

Apply **data analysis, machine learning, and visualization** to uncover actionable insights, enhance customer experience, and guide brand strategies.

Contents

To analyze the dataset and derive reliable insights, we first perform data cleaning and exploratory data analysis, followed by applying various analytical techniques.

Data Overview - Dataset & Key Features

Product Information Dataset

- **8,000+ products** across various categories.
- **Key features:**
  - Product name, brand, category, price, ingredients, average ratings, and attributes (size, online_only, etc.)
- **Purpose:**
  - Provide detailed product information, including pricing, ingredients, and categories, while providing an overview of customer satisfaction.

Customer Reviews Dataset

- **1 million+ reviews** on **2,000 skincare products.**
- **Key features:**
  - Ratings, review date, review text, and customer information (skin type, eye color, etc.)
- **Purpose:**
  - Provide detailed user feedback with temporal information to enable deeper insights into customer satisfaction.

Data Overview - Preprocessing Steps

1

Data Cleaning

- **Missing Values:** Null values in product attributes and customer information were addressed through **reasonable imputation or removal**.
- **Data Manipulation**: Extracted numerical data from text-based fields

2

Data Integration

Merged the product information dataset with customer review datasets to establish links between product attributes and customer feedback.

3

Feature Transformation

- **Numerical Features**: Applied **log transformation** to normalize exponential distributions.
- **Categorical Features**: Used **one-hot encoding** to convert categories into binary vectors.
- **Text Data**: Implemented **TF-IDF encoding** to quantify text relevance.

4

Dimension Reduction

Applied **Singular Value Decomposition** to reduce the dimensionality of **text features and sparse categorical features**

Analytical Methods

Numerical Relationship - Correlation Models

- Regression Analysis
- ANOVA
- Pivot Table
- Chi-Squared, T-Testing
- etc.

Graphical Analysis - Visualization Tools

- Heatmap
- Violin Plot
- Hexbin Plot
- Barchart
- etc.

Classification - Machine Learning Models

- SVM
- XGBoost
- Random Forest
- Gradient Boosting
- Voting Classifier
- etc.

KEY FINDINGS

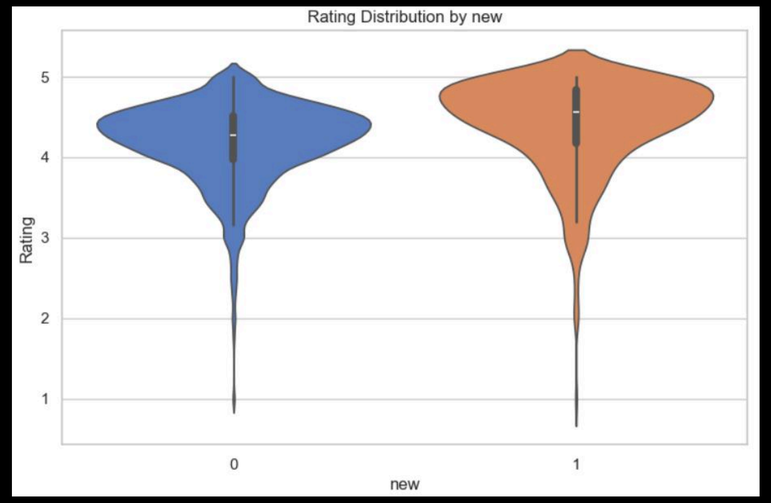
Factors driving high product ratings

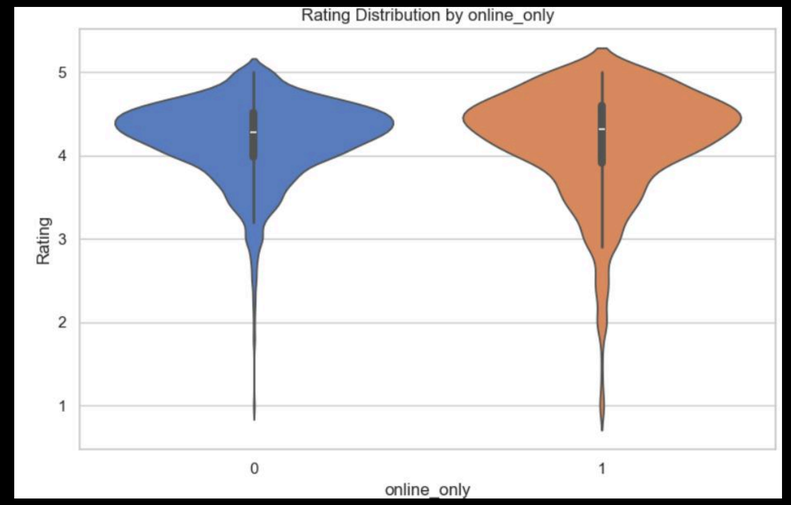
Related Boolean Feature:

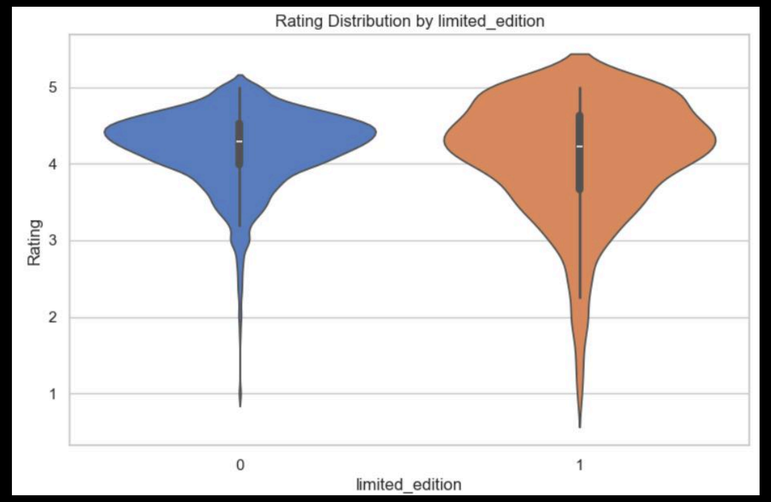
We use **ANOVA** and **violin plot** to investigate how these factors are related to high product ratings

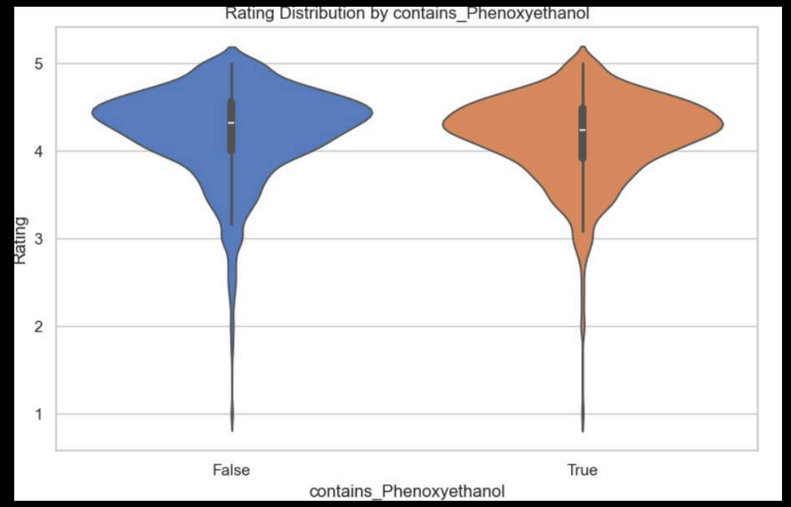
Related Factors:

- Online Only
- New
- Limited Edition
- Specific Ingredients including: Phenoxyethanol, Tocopherol, Limonene, Linalool, Silica, Butylene Glycol, etc.









KEY FINDINGS

Factors driving high product ratings

Key Insights

We calculated the mean product rating for different brand and category.

Related Factors:

- Brand
- Primary Category
- Secondary Category
- Tertiary Category
- Interaction Terms:
  - Primary Category & Online-only
  - Secondary Category & Online-only
  - Tertiary Category & Online-only

Challenge: How to manage sparse categorical features and their interactions.

Category v.s. Rating

	primary_category	average_rating
2	Gifts	4.563450
5	Men	4.504992
8	Tools & Brushes	4.271458
1	Fragrance	4.230889
7	Skincare	4.228890
3	Hair	4.201113
0	Bath & Body	4.195015
4	Makeup	4.146845
6	Mini Size	4.005665

Brand vs Average Rating

	brand_name	average_rating
78	Erno Laszlo	5.000000
10	Aquis	4.904800
158	MACRENE actives	4.889420
161	MARA	4.823860
32	CANOPY	4.813733
..	...	...
47	Christophe Robin	3.154767
101	Good Dye Young	3.062050
194	Overose	2.848450
253	The Maker	NaN
299	philosophy	NaN

Rating v.s. Online or not & Category

primary_category	Non-Online	Online	Rating_Difference
Mini Size	3.998237	4.058100	0.059863
Skincare	4.218118	4.271438	0.053320
Tools & Brushes	4.271543	4.271013	-0.000530
Hair	4.205973	4.191046	-0.014927
Bath & Body	4.222076	4.151288	-0.070788
Fragrance	4.259239	4.144556	-0.114683
Men	4.542989	4.382857	-0.160132
Makeup	4.164505	3.998730	-0.165774
Gifts	4.563450	NaN	NaN

Rating v.s. Online or not & Category

pivot_table top 10:

tertiary_category	Non-Online	Online	Rating_Difference
Teeth Whitening	3.239400	4.462040	1.222640
Blotting Papers	3.823067	5.000000	1.176933
Cologne Gift Sets	4.187500	5.000000	0.812500
Curling Irons	3.864867	4.568389	0.703522
Hair Thinning & Hair Loss	3.917200	4.544000	0.626800
Tinted Moisturizer	4.012629	4.418500	0.405871
Sheet Masks	4.080966	4.486100	0.405134
Eye Brushes	4.476147	4.857100	0.380953
Concealer	4.205592	4.541175	0.335583
Eye Masks	3.991108	4.307500	0.316392

pivot_table last 10:

tertiary_category	Non-Online	Online	Rating_Difference
Brush Cleaners	4.390150	3.818067	-0.572083
Eyelash Curlers	4.001762	3.379800	-0.621962
Face Sets	4.306533	3.653525	-0.653008
Hair Dye & Root Touch-Ups	4.311633	3.491200	-0.820433
False Eyelashes	4.092528	3.270025	-0.822503
Lip Gloss	4.240900	3.411933	-0.828967
Sponges & Applicators	4.198200	3.275133	-0.923067
Bath Soaks & Bubble Bath	4.066560	2.802650	-1.263910
Makeup Bags & Travel Cases	4.946650	3.666700	-1.279950
Lip Plumper	4.066493	1.000000	-3.066493

KEY FINDINGS

Factors driving high product ratings

Ratings by Category:

- High Ratings:
  - Gifts, Men's (well-met needs)
- Low Ratings:
  - Mini Size, Makeup (value issues);

Online vs. Offline Impact:

- Perform Better Online:
  - Skincare, Beauty Tools, Teeth Whitening (detailed reviews).
- Perform Better Offline:
  - Fragrance, Makeup (sensory testing).

Category v.s. Rating

	primary_category	average_rating
2	Gifts	4.563450
5	Men	4.504992
8	Tools & Brushes	4.271458
1	Fragrance	4.230889
7	Skincare	4.228890
3	Hair	4.201113
0	Bath & Body	4.195015
4	Makeup	4.146845
6	Mini Size	4.005665

Brand vs Average Rating

	brand_name	average_rating
78	Erno Laszlo	5.000000
10	Aquis	4.904800
158	MACRENE actives	4.889420
161	MARA	4.823860
32	CANOPY	4.813733
..	...	...
47	Christophe Robin	3.154767
101	Good Dye Young	3.062050
194	Overose	2.848450
253	The Maker	NaN
299	philosophy	NaN

Rating v.s. Online or not & Category

primary_category	Non-Online	Online	Rating_Difference
Mini Size	3.998237	4.058100	0.059863
Skincare	4.218118	4.271438	0.053320
Tools & Brushes	4.271543	4.271013	-0.000530
Hair	4.205973	4.191046	-0.014927
Bath & Body	4.222076	4.151288	-0.070788
Fragrance	4.259239	4.144556	-0.114683
Men	4.542989	4.382857	-0.160132
Makeup	4.164505	3.998730	-0.165774
Gifts	4.563450	NaN	NaN

Rating v.s. Online or not & Category

pivot_table top 10:

tertiary_category	Non-Online	Online	Rating_Difference
Teeth Whitening	3.239400	4.462040	1.222640
Blotting Papers	3.823067	5.000000	1.176933
Cologne Gift Sets	4.187500	5.000000	0.812500
Curling Irons	3.864867	4.568389	0.703522
Hair Thinning & Hair Loss	3.917200	4.544000	0.626800
Tinted Moisturizer	4.012629	4.418500	0.405871
Sheet Masks	4.080966	4.486100	0.405134
Eye Brushes	4.476147	4.857100	0.380953
Concealer	4.205592	4.541175	0.335583
Eye Masks	3.991108	4.307500	0.316392

pivot_table last 10:

tertiary_category	Non-Online	Online	Rating_Difference
Brush Cleaners	4.390150	3.818067	-0.572083
Eyelash Curlers	4.001762	3.379800	-0.621962
Face Sets	4.306533	3.653525	-0.653008
Hair Dye & Root Touch-Ups	4.311633	3.491200	-0.820433
False Eyelashes	4.092528	3.270025	-0.822503
Lip Gloss	4.240900	3.411933	-0.828967
Sponges & Applicators	4.198200	3.275133	-0.923067
Bath Soaks & Bubble Bath	4.066560	2.802650	-1.263910
Makeup Bags & Travel Cases	4.946650	3.666700	-1.279950
Lip Plumper	4.066493	1.000000	-3.066493

KEY FINDINGS

Time Trends of Customer Sentiment by Category

Stable Categories:

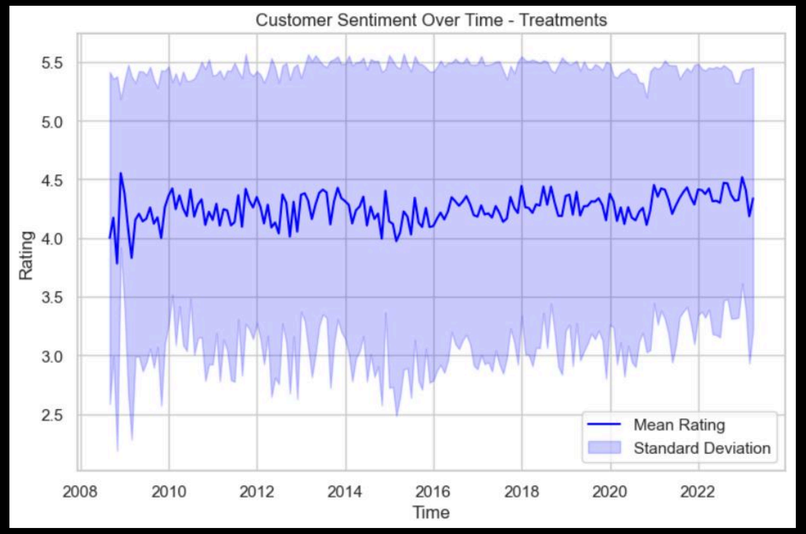
- Eye Care, Treatments
- Consistent ratings → Maintain quality & gather feedback.

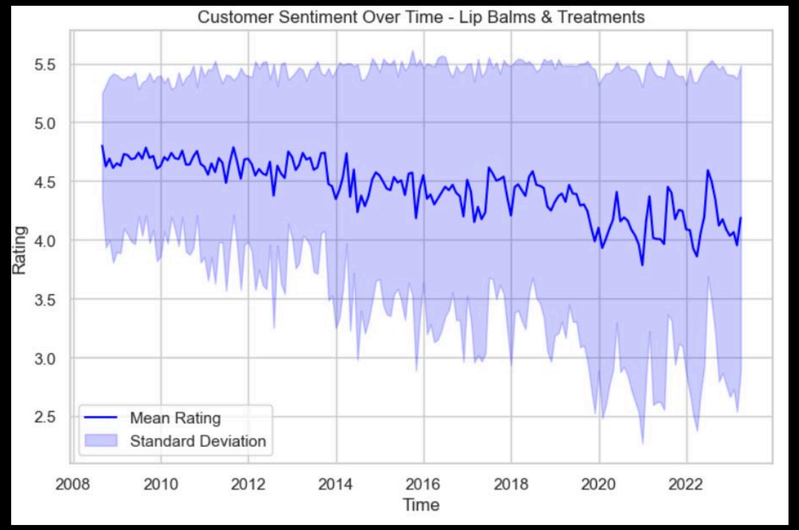
Declining Categories:

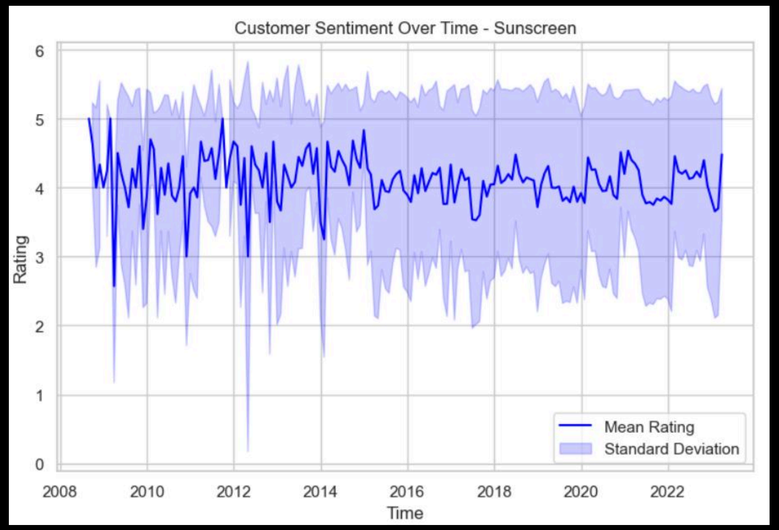
- Lip Balms & Treatments
- Ratings decline → Investigate issues & improve/reposition products.

Fluctuating but Stabilizing Categories:

- Sunscreen, High Tech Tools
- Stabilizing ratings → Ensure consistent quality & targeted marketing to build loyalty.







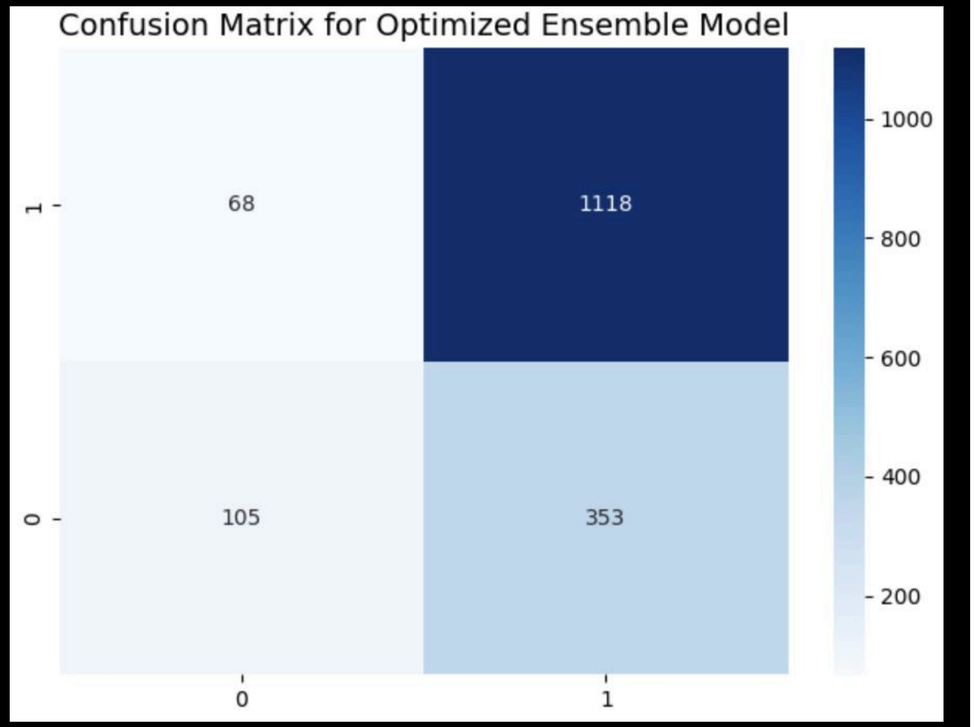
KEY FINDINGS

Machine Learning Models to Predict Product Rating

We predict product ratings using features (e.g., product name, brand, category, price, ingredients, and other attributes and identify key drivers of customer satisfaction.

Final Choice of Model:
Voting Classifier

- Support Vector Machine
- Random Forest
- Gradient Boosting
- XGBoost



ACCURACY

0.7439

PRECISION

0.7600

RECALL

0.9427

F1 SCORE

0.8416

AUC

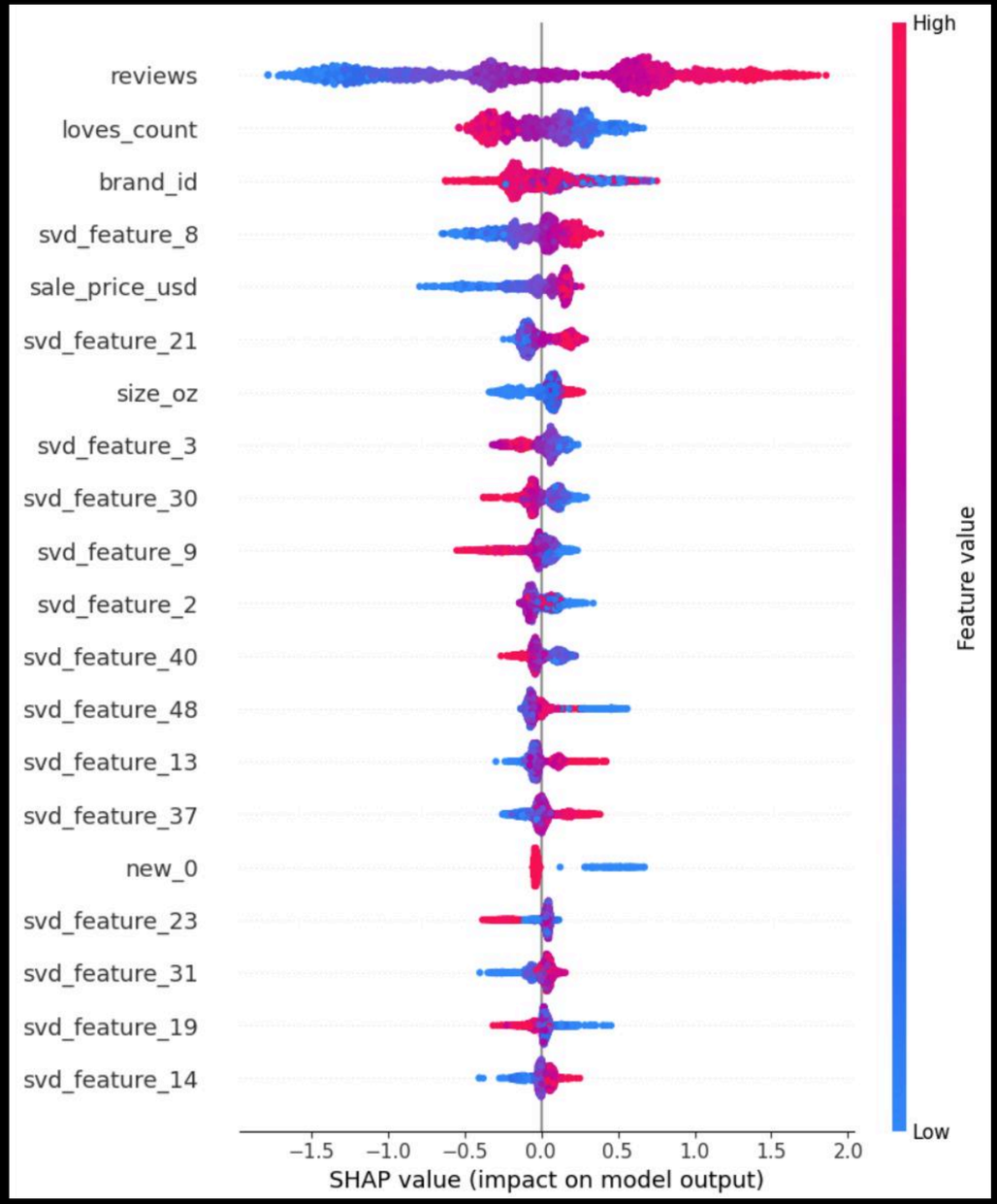
0.7654

KEY FINDINGS

Machine Learning Models to Predict Product Rating

According to the **SHAP analysis**, the following features significantly contribute to higher average product ratings:

- Reviews
- Loves_count
- Brand_id
- Size_oz
- SVD Features:
  - Product_name
  - Highlights
  - Secondary_category
  - Tertiary_category
  - Ingredients
- Other relevant features



Conclusions & Implications

Conclusions

- Product ratings are influenced by various features, and it is possible to make approximate predictions based on certain attributes.
- Products from different categories may be better suited for either online or offline scenarios

Implications

- Brand Strategy: Take different online & offline strategy according to its category.
- Product Design: Model and correlation results provide insights into how to better refine on product design

Teamwork and Participation

- Ruizhe Wang (wrzwrz): Statistical Analysis, Model Fine-tuning, Model Interpretation, Graphical Analysis

- Zhiyi Ji (jizhiyi): Introduction, Data Overview, Data Preprocessing, Statistical Analysis, Model Implementation, Graphical Analysis

Credit

- Nedap Retail's case study on Sephora: https://www.nedap-retail.com/case-study-sephora/
- Kaggle dataset "Sephora Products and Skincare Reviews" by Nady Inky: https://www.kaggle.com/datasets/nadyinky/sephora-products-and-skincare-reviews

PRODUCT

THANK YOU

Group 23
Ruizhe Wang, Zhiyi Ji
12/06/2024