

Problem Statement

- Highly Competitive Industry
- Challenge: continuously satisfying customers with a wide range of product choice available
- Significance:
 - Drive product innovation
 - Foster customer loyalty
 - Strengthen Sephora's market position



Dataset Description

- Dataset:
 - Products:
 - **8,494** rows and **27** columns
 - Reviews:
 - **5 tables** (merged into 1 table)
 - **1 million** rows and **19** columns
- Data gathered from **2008-2022**

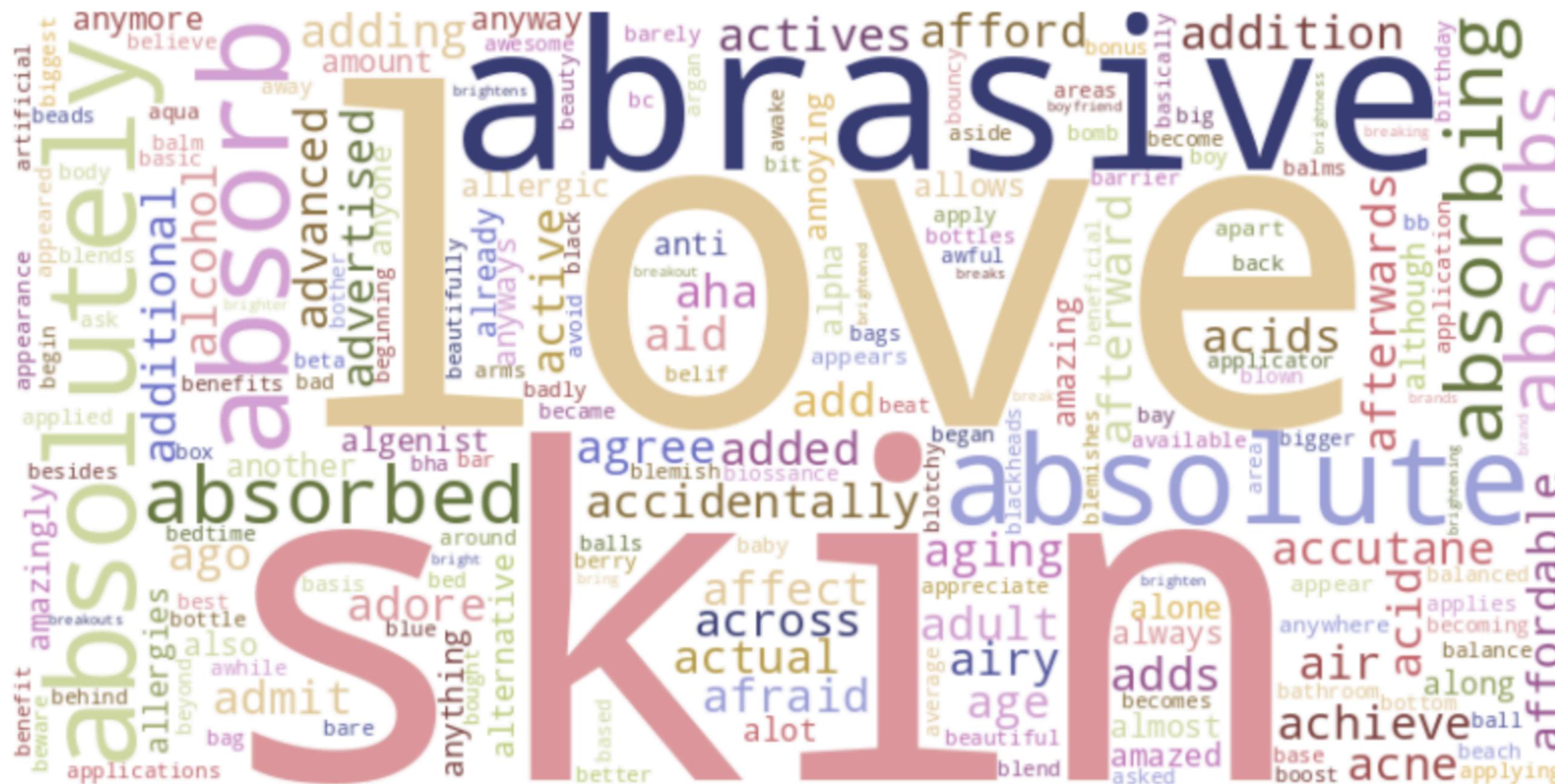


Market Basket Analysis

Entry	Antecedents	Consequents	Support	Confidence	Lift
3	'Green Clean Makeup Meltaway Cleansing Balm'	['Green Clean Makeup Removing Cleansing Balm']	0.0054	0.9959	140.45
14	'Mini Superfood Antioxidant Cleanser'	['Superfood Antioxidant Cleanser']	0.0049	0.9879	149.58
1	'Daily Microfoliant Exfoliator'	['Mini Daily Microfoliant Exfoliator']	0.0046	0.9939	188.53
28	'(Set, Glow Face Mist)'	['Beauty Elixir Prep']	0.0034	0.9785	256.99

- Normal size sees association with its mini version - potentially due to different sizes of products get the same review
- Similar product types see association

TF-IDF on Reviews



The word cloud showcases words with high TF-IDF: a mix of positive and negative words

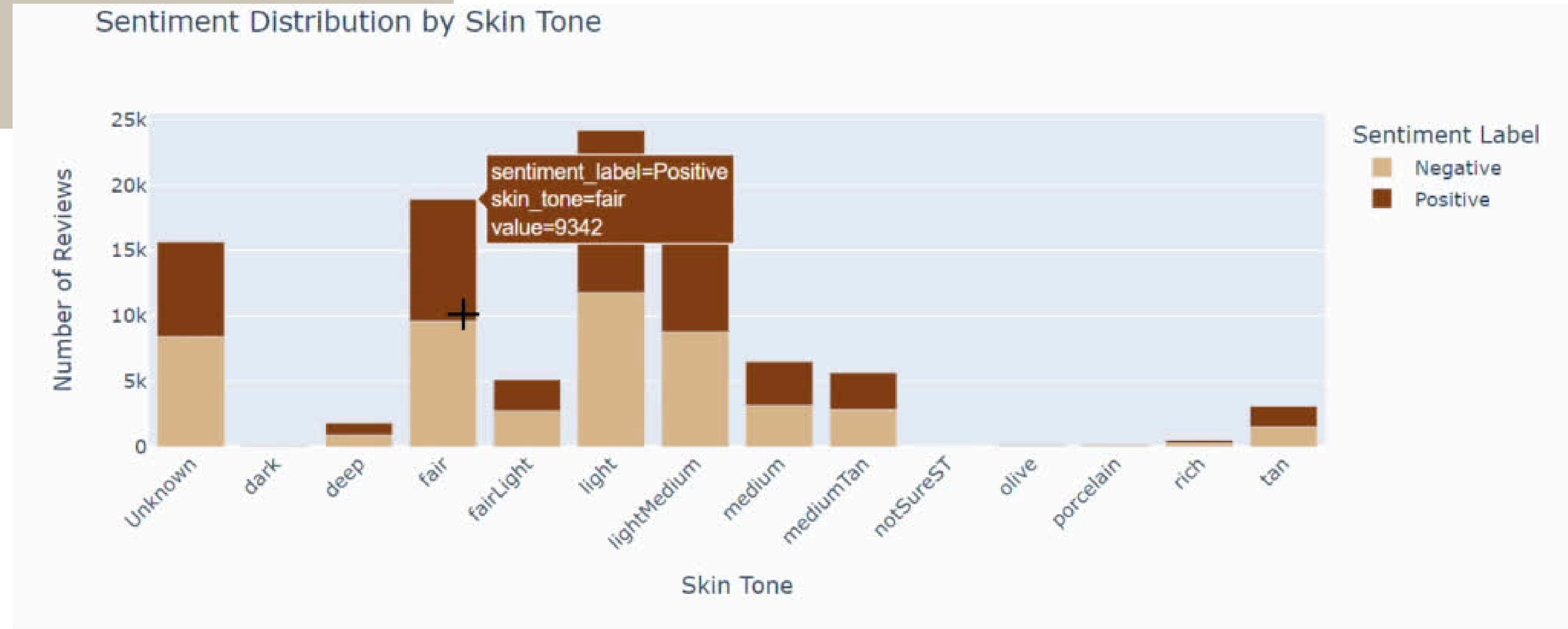
Positive Reviews Wordcloud



Negative Reviews Wordcloud



Sentiment Score by Skin Tone

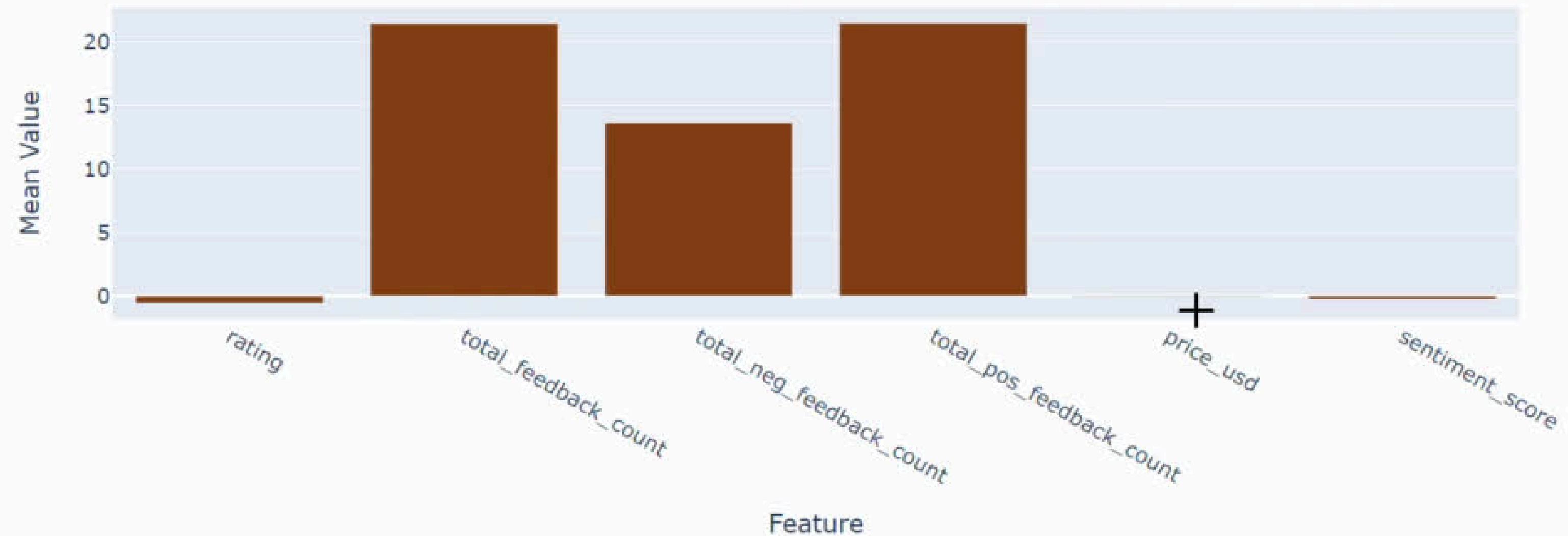


Sentiment Analysis

- Analyzed **customer reviews** sentiment for Sephora products
- Skincare products show **polarized opinions**
- Sephora addresses **diverse customer needs**, including varying skin tones and types
- Customers have **strongly positive or negative experiences** across all skin features

Clustering Feature Importance

Mean Feature Values Across Clusters

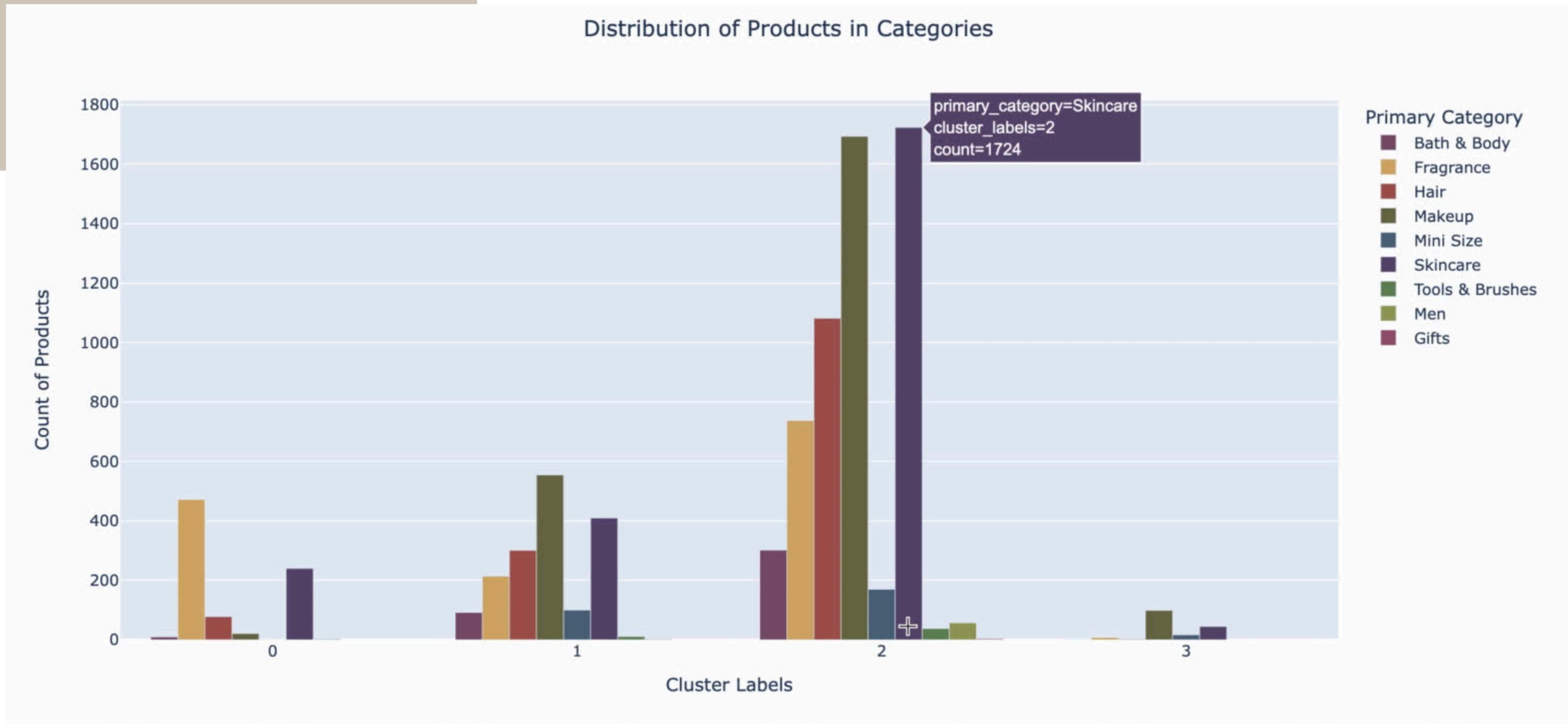


Clustering on Reviews

- Conducted KMeans clustering on review variables
- Feedback count emerged as most influential
- Sentiment scores had minimal impact
- Price significant for high-priced products
- **Suggestion:** Focus on **high-priced** items for attention due to their impact



Clustering Products



Challenges

- Limited computational power required downsizing the sample from 1 million to 100,000
- Dilemma with **mini** and **full sizes** of products, yielding similar reviews due to being the same product
- Challenges encountered with **GitHub** collaboration
- Incompatibility of Vertex AI with **3D/interactive graphs**

Business Recommendations

- Make skin tones more inclusive
- Bundle products within the same category
- Focus on skincare product line
- Analyze clustering results to identify high-performing categories
- Enhance customer review engagement
- Promote products that see high positive score



**Thank you!
Any questions?**