Sephora Reviews
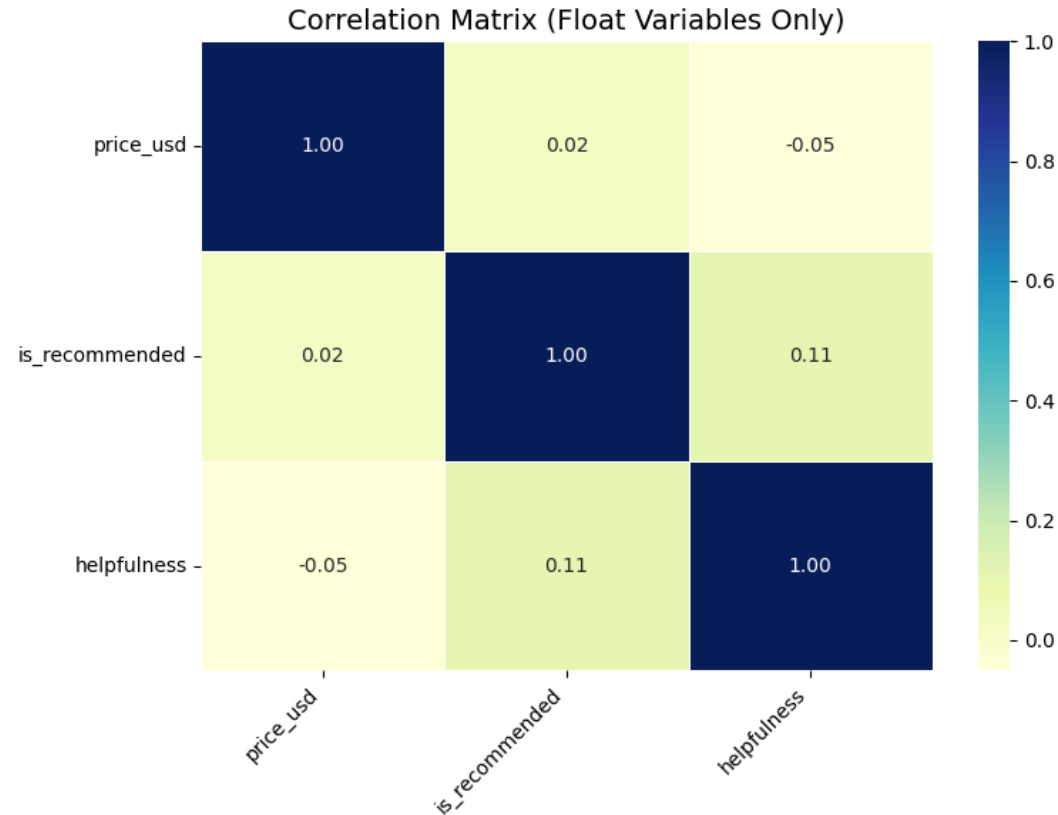
Le Thu Trang Truong

- **Data Cleansing & Brand Integrity**
- **Customer Segmentation & Product Optimization**

Summary from Reviews

Correlation Matrix (Float Variables Only)

Helpful reviews influence recommendations more than price.

Product Filtering Criteria

|	avg_rating	review_count	recommended_count	not_recommended_count
count	1597.000000	1597.000000	1597.000000	1597.000000
mean	4.201848	106.007514	82.867877	23.139637
std	0.522349	103.040913	88.356255	36.815021
min	1.000000	1.000000	0.000000	0.000000
25%	3.933333	19.000000	13.000000	3.000000
50%	4.287129	69.000000	46.000000	10.000000
75%	4.573864	172.000000	133.000000	28.000000
max	5.000000	492.000000	370.000000	315.000000

Remove?

1. Very few reviews
- Fewer than 10 reviews
2. Low engagement & negative sentiment
- Between 10–18 reviews, and:
  - Average rating ≤ 2
  - Not recommended ratio ≥ 75%

- Products to remove: 224
- Products to keep: 1373

Customer Segmentation Product Optimization

Optimal Number of Topics
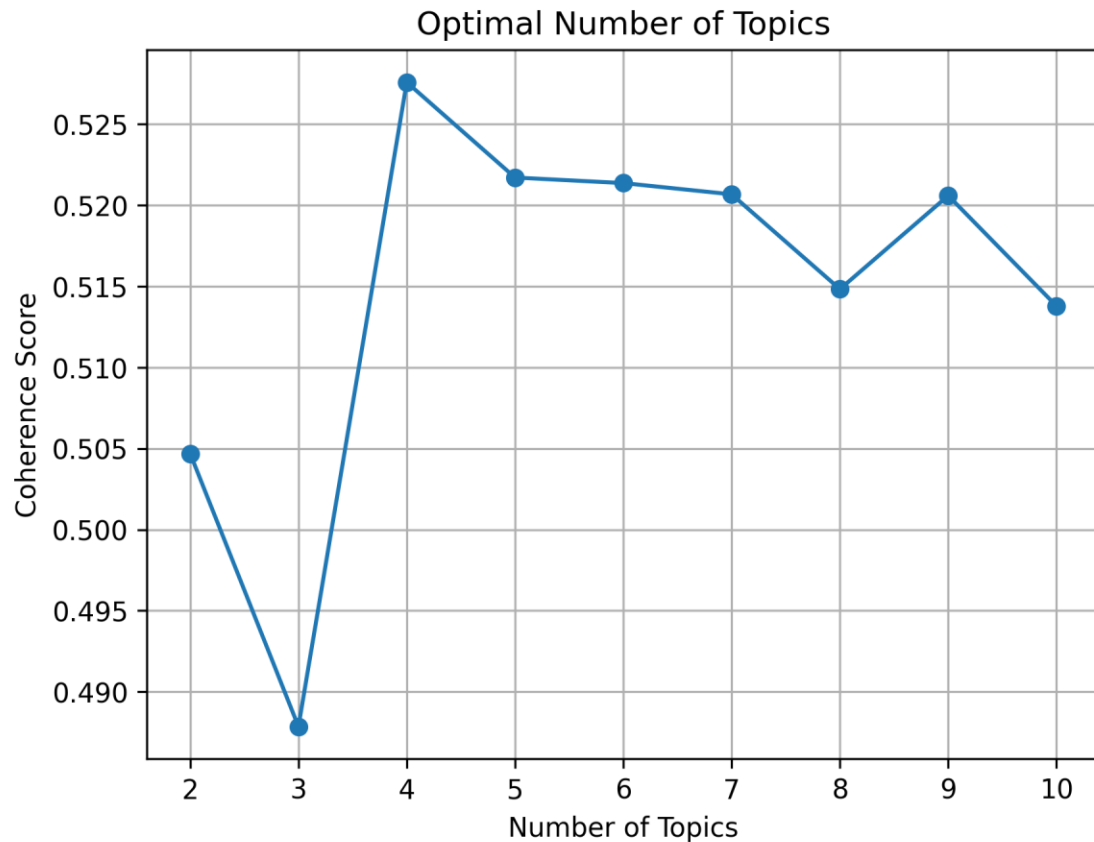
- **Topic coherence** measures how meaningful and consistent the topics are.
- The highest coherence score was reached with **4 topics**.
- Indicates that 4-topic LDA produces the most interpretable themes from negative reviews.

Topic	Image	Content
1		**Luxury vs Practicality Clash** • Frequent words: collect, dense, elegance, contrast, circulating • Insight: Users may feel the product focuses too much on aesthetics or luxury appeal, while lacking in real performance or practicality.
2		**Texture & Irritation Issues** • Frequent words: edition, deodorant, dense, bamboo, flaky • Insight: Complaints suggest flaky or irritating consistency, especially in deodorants or scented products, leading to dissatisfaction.
3		**Unnatural Texture & Confusion** • Frequent words: edition, collect, dense, behavior, bamboo • Insight: Customers mention confusing or inconsistent texture, possibly tied to limited edition or dense formulations that don't blend well.
4		**Color & Skin Reaction Concerns** • Frequent words: attracted, bluish, complaint, blush, collect • Insight: Some customers report strange coloration or unexpected skin reactions like a bluish tint, indicating sensitivity or mismatched expectations.

NLP-Driven Recommendation Rule Engine

index	Brand	Product	Ideal Customer
0	algenist	10 day results kit	Normal/oily skin, non-reactive users
1	algenist	aa (alguronic acid) barrier serum	Normal skin, users seeking hydration
2	algenist	advanced anti-aging repairing oil	Users applying away from eyes
3	algenist	algae niacinamide moisture veil	Normal/oily skin, non-reactive users
4	algenist	alive prebiotic balancing mask	Normal/oily skin, non-reactive users
5	algenist	alive prebiotic balancing moisturizer spf 15	Normal/oily skin, non-reactive users
6	algenist	blue algae vitamin c dark spot correcting peel	Normal/oily skin, non-reactive users
7	algenist	blue algae vitamin c skinclarity brightening serum	Normal/oily skin, non-reactive users
8	algenist	elevate advanced lift contouring cream	Users applying away from eyes
9	algenist	elevate advanced retinol serum	Normal/oily skin, non-reactive users

Specific skincare products were mapped to their ideal customer profiles based on NLP-driven topic modeling of low-rated reviews.

Conclusion

Full implementation available on GitHub

- Used **SQL** to filter, join, and prepare structured review data for analysis.
- Applied **NLP topic modeling (LDA)** to uncover key reasons behind **low product ratings**.
- Built a **Rule-Based Recommendation Engine** to match products with ideal customer profiles.
- Enabled targeted product positioning by **predicting preferences** based on reviews and ratings over time.

Result: A clearer path to personalized skincare recommendations and better customer satisfaction.