

From Reviews to Insights: A 3-Stage Pipeline for Sephora Customer Sentiment and Insights

6 Nov 2025

Group 5:
Chen Peng Wei
Deepika Raja
Ding Hong-Yan
Kim Jeounghyun
Moo Jia Rong
Tan Zong Yin Shawn
Xie Haoyang

Background & Motivations

Context:

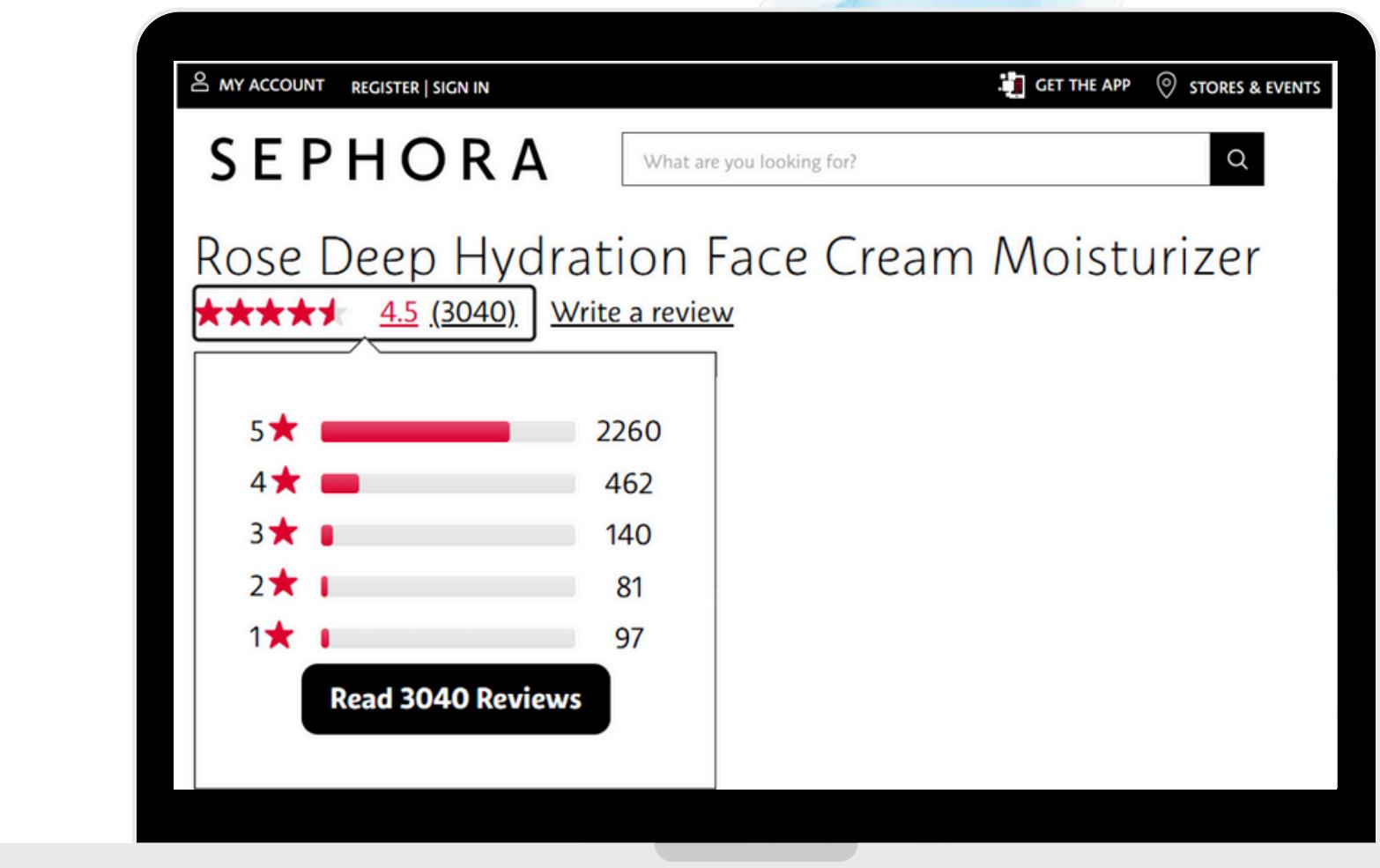
- Sephora's e-commerce platform with massive volumes of valuable, unstructured customer reviews.

Challenge:

- Manual analysis is infeasible at scale.
- Internal teams struggle to extract actionable insights for product strategy and marketing.
- Customers face difficulty distilling meaningful information for purchase decisions.

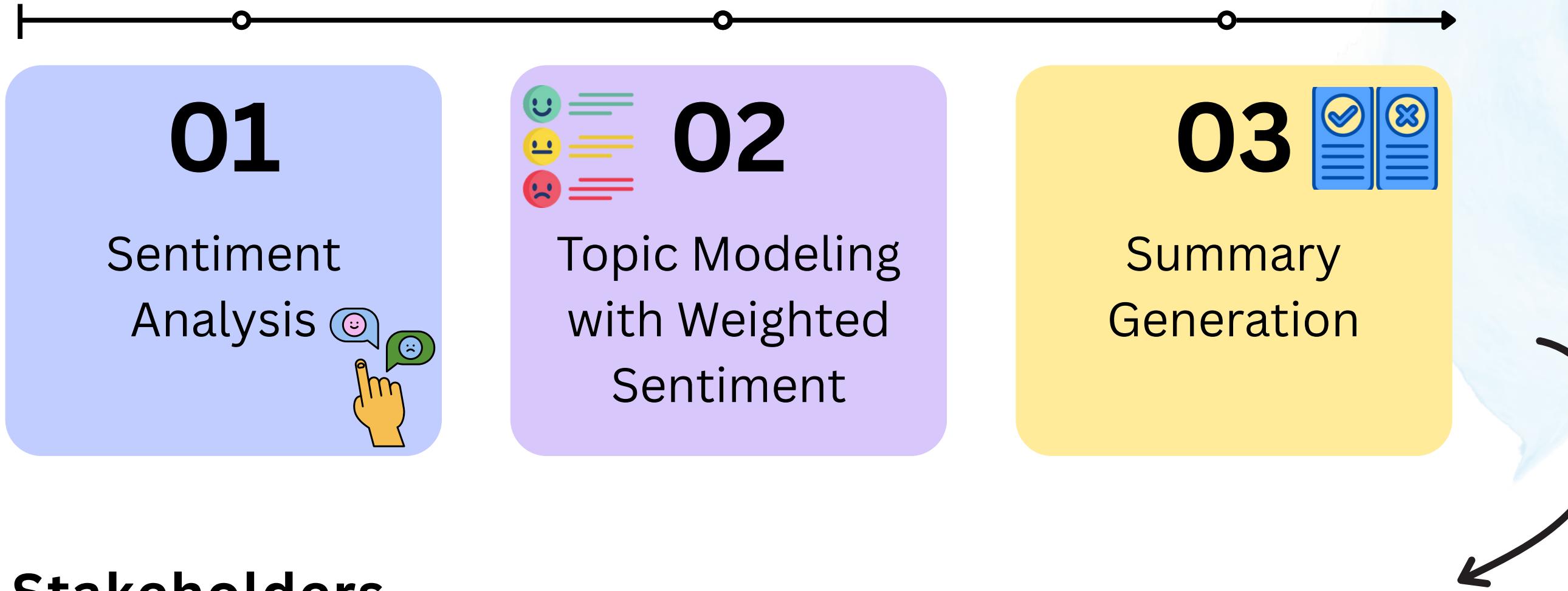
Impact:

- Missed opportunities for product improvement.
- Risk of misallocated marketing resources.
- Potential negative effect on customer satisfaction and loyalty.



Project Objectives

3-stage pipeline to transform raw, unstructured data into actionable intelligence:



Stakeholders

1. Internal Teams

- Empowers product, marketing, and customer experience teams with data-driven insights
- Enables rapid detection of emerging product issues
- Supports targeted, evidence-based marketing strategies

2. Customers

- Provides clear, summarized feedback insights
- Enhances confidence and efficiency in purchase decisions
- Improves overall customer experience

Introduction of Dataset

Source:

- **Sephora Products and Skincare Reviews From Kaggle**, collected via Python web scraping from Sephora online store in March 2023

Data Scale:

- 8,000+ beauty products across multiple categories
- 1,000,000+ user reviews on 2,000+ skincare products

Scope:

- Includes product information and user reviews

1. Product Data

Contains product-level information such as:

- Product & brand name, price, ingredients, size, variation
- Edition status (new/limited), stock status, exclusivity
- Primary & secondary categories

2. Reviews Data

Contains review-level information such as:

- Author ID, rating, recommendation flag, review text & title
- Feedback metrics (helpfulness, positive/negative feedback counts)
- User appearance attributes (skin type, tone, eye color, hair color)
- Submission time



Review Selection Process

Step 1: Category level

- Selected top 3 categories – Moisturizers, Treatments, Cleansers – with sufficient reviews for reliable NLP analysis.

Step 2: Brand level

- Chose top 3 brands per category based on review volume to ensure diversity.

Step 3: Product level

- Picked top 5 products per brand (Combined variants under main product e.g., “Mini Size”, “Travel Set”) and excluded products with fewer than 400 reviews.

Step 4: Review level:

- 400 reviews per product – for balanced and diverse sentiment coverage.

The Results

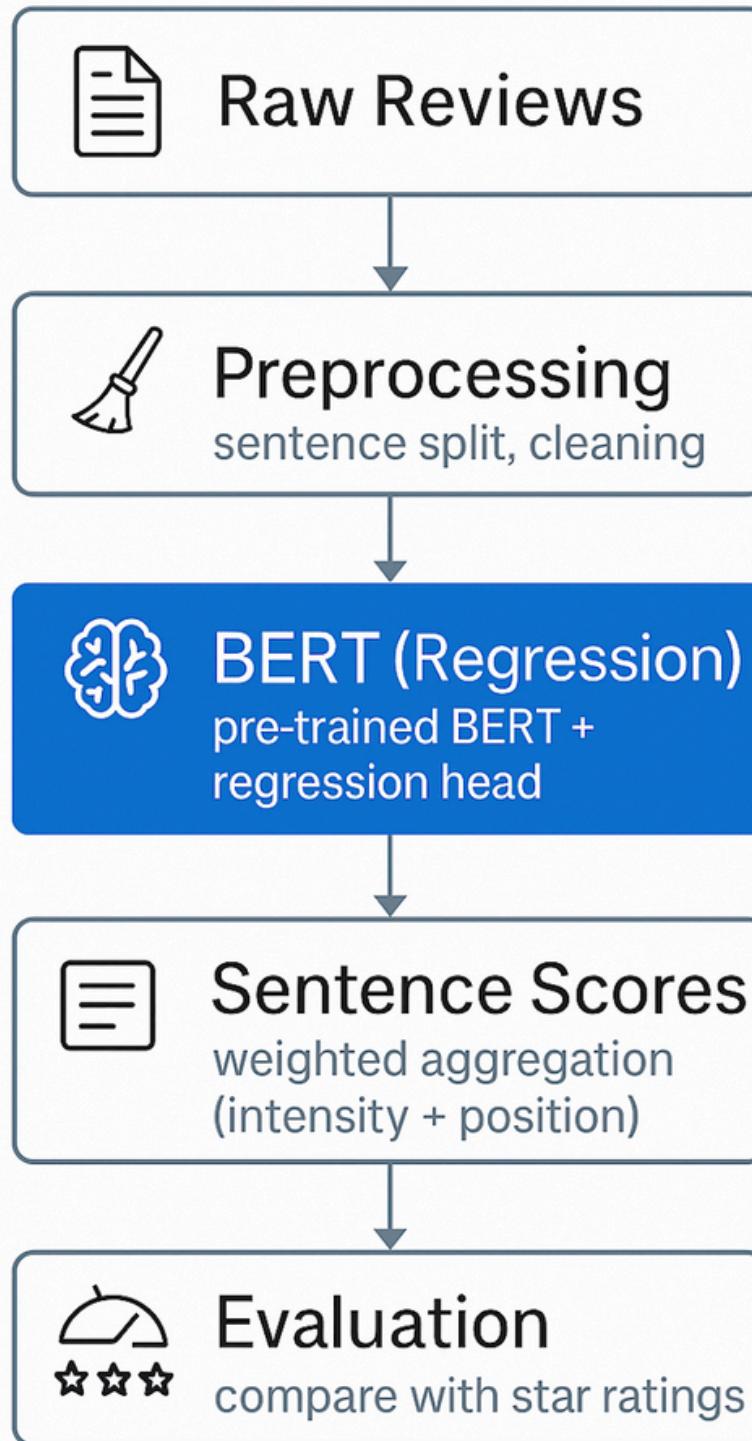
- 3 Categories
- 9 Brands
- 42 Products
- 16800 Reviews



Sentiment Analysis

Result & Examples

Task 1—Sentiment Analysis Workflow



review_seq_id	Sentence Text	Sentiment Score
690007	"Terrible product!"	-0.95
690007	"The foam falls right off your hand as soon as it touches water."	-0.82
690007	"It's so foamy that none of it gets on your face."	-0.75
690007	"Waste of money."	-0.91

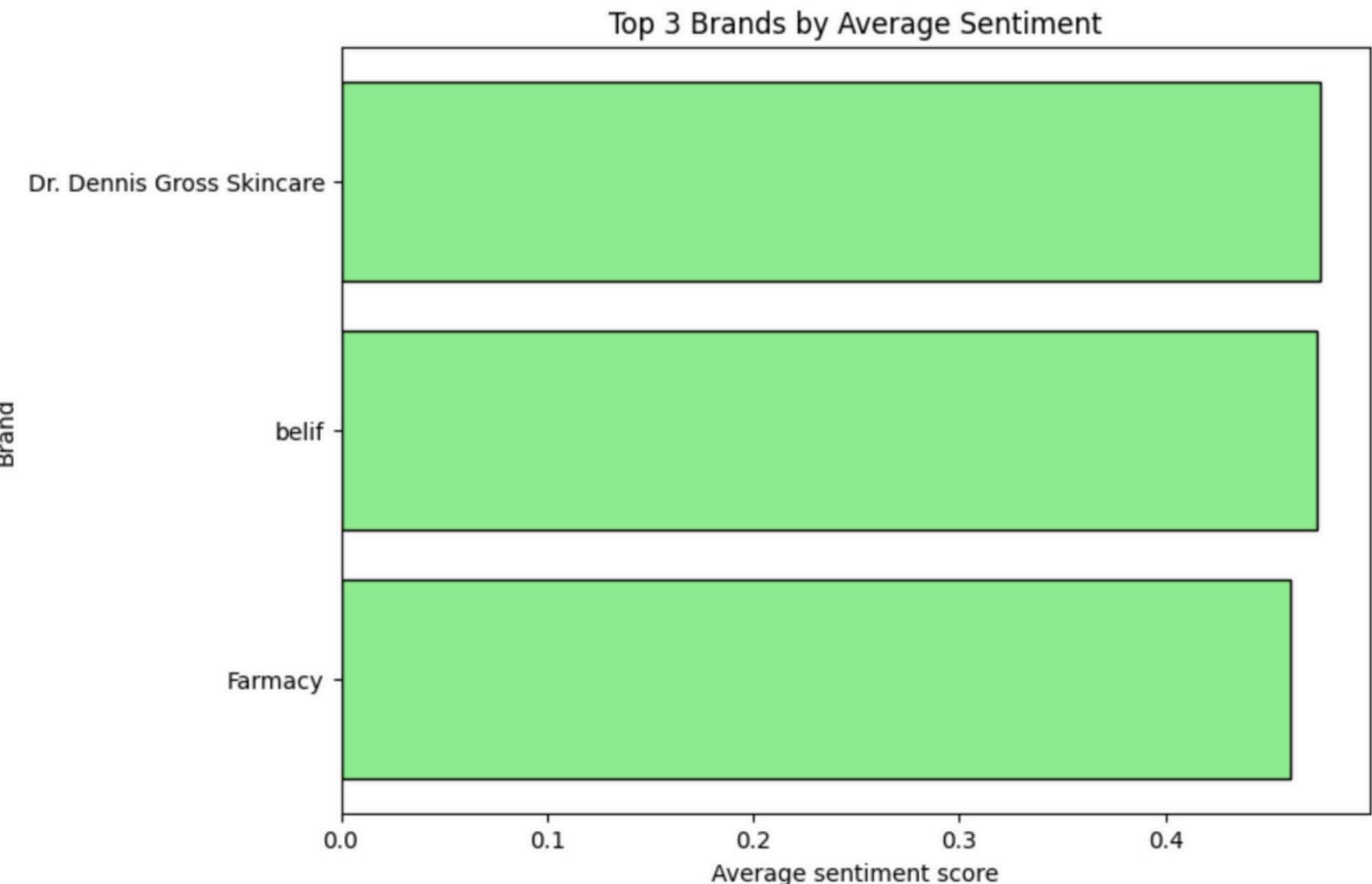
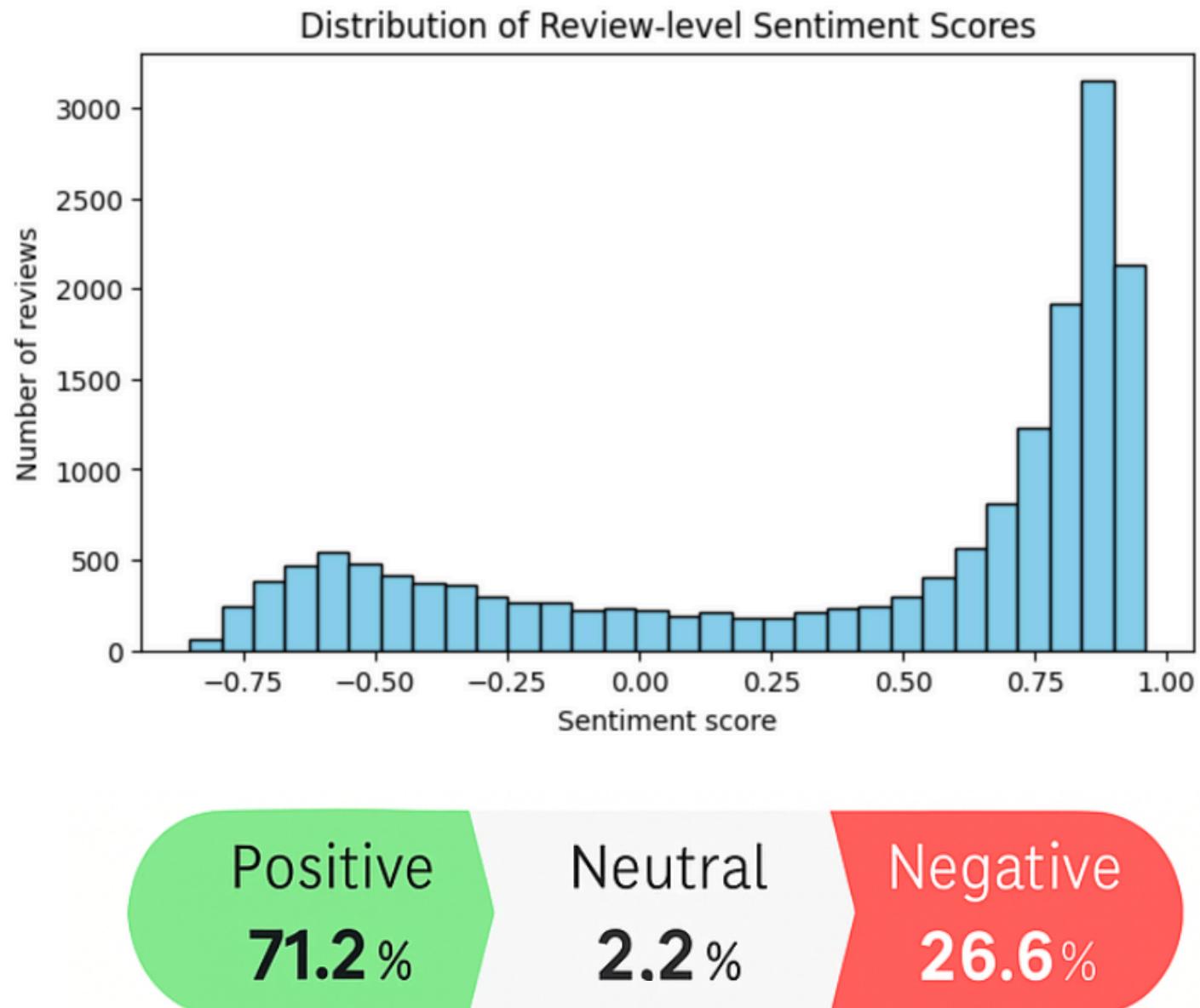
review_seq_id	Aggregated Review Sentiment	Star Rating
690007	-0.85	★ (1)

Model Evaluation: Confusion Matrix & Performance Metrics

True Star (Actual) →	Pred 1	Pred 2	Pred 3	Pred 4	Pred 5
★ 1 (actual 1 star)	905	703	221	10	3
★★ 2	300	936	525	36	8
★★★ 3	50	481	1133	324	89
★★★★ 4	1	21	310	749	1901
★★★★★ 5	0	1	42	259	7792

Metric	Value
Mismatch Rate	0.315
Pearson Correlation	0.914
Spearman Correlation	0.847
MAE / RMSE	0.468 / 0.591

Sentiment Analysis



Most reviews show strong positive sentiment, with many customers expressing emotions intensely—reflecting **high satisfaction and engagement**.

Mismatch Analysis Between Model Predictions & User Ratings

Process:

- Converted model sentiment scores into star ratings and compared them with user-provided ratings to assess model accuracy and robustness.

Focus:

- Analyzed cases with a ≥ 2 -star difference between predicted and actual ratings – indicating major sentiment misinterpretations.

Outcome:

- Identified two main types of mismatches in model prediction.

Star Review	Sentiment Score Range	Interpretation
★ (1 star)	[-1.0 ~ -0.6]	Extremely Negative
★★ (2 stars)	(-0.6 ~ -0.3]	Moderately Negative
★★★ (3 stars)	(-0.3 ~ 0.3)	Neutral
★★★★ (4 stars)	[+0.3 ~ +0.6]	Moderately Positive
★★★★★ (5 stars)	(+0.6 ~ +1.0]	Extremely Positive

1. Model-Related Limitations

Type	Description	Example (User ★ / Model ★)	Key Insight
(a) Mixed Sentiments	Model fails to interpret reviews containing both praise and criticism.	<p>“Works like a charm taking stubborn eye make up off. Im almost out of product. Makes me sad :(”</p> <p>User: 5★ / Model: 3★</p>	Model averages conflicting emotions instead of detecting tone shifts.
(b) Misinterpreting Personal Sensitivity	Negative words about skin sensitivity are wrongly taken as product dissatisfaction.	<p>“Did a lovely job exfoliating, but the second step burned... Too strong for my sensitive skin.”</p> <p>User: 4★ / Model: 1★</p>	Model overlooks that product is effective but personally unsuitable.
(c) Sarcasm Detection Failure	Model misreads sarcasm as positive sentiment.	<p>“Awesome. I wasted \$90 for nothing.”</p> <p>User: 1★ / Model: 4★</p>	Model fails to recognize sarcastic tone indicating frustration.

2. Review-System-Related Mismatches

Type	Description	Example (User ★ / Model ★)	Key Insight
(a) Accidental Misrating	Users may click the wrong rating by mistake, causing a mismatch between text and stars.	"I have had this as a sample... makes my face feel so soft, in love." User: 1★ / Model: 5★	Rating error likely due to misclick or confusion , not sentiment inconsistency.
(b) System-Induced Rating	Platform forces users to choose a star before posting a question or comment.	"Jus a question... I only put yes, n a star because it wouldn't allow me to leave the page w/o answering lol." User: 1★ / Model: 5★	Low rating stems from system design , not user opinion; text expresses curiosity, not dissatisfaction.

Recommendations

- **Intelligent Validation System**

- Analyze review text sentiment before submission.
- Trigger a smart warning when a large mismatch is detected:
 - “⚠ Warning: Your text appears strongly positive, but you selected 1 star. Is this correct?”
- Helps prevent accidental or inconsistent ratings, improving feedback reliability.

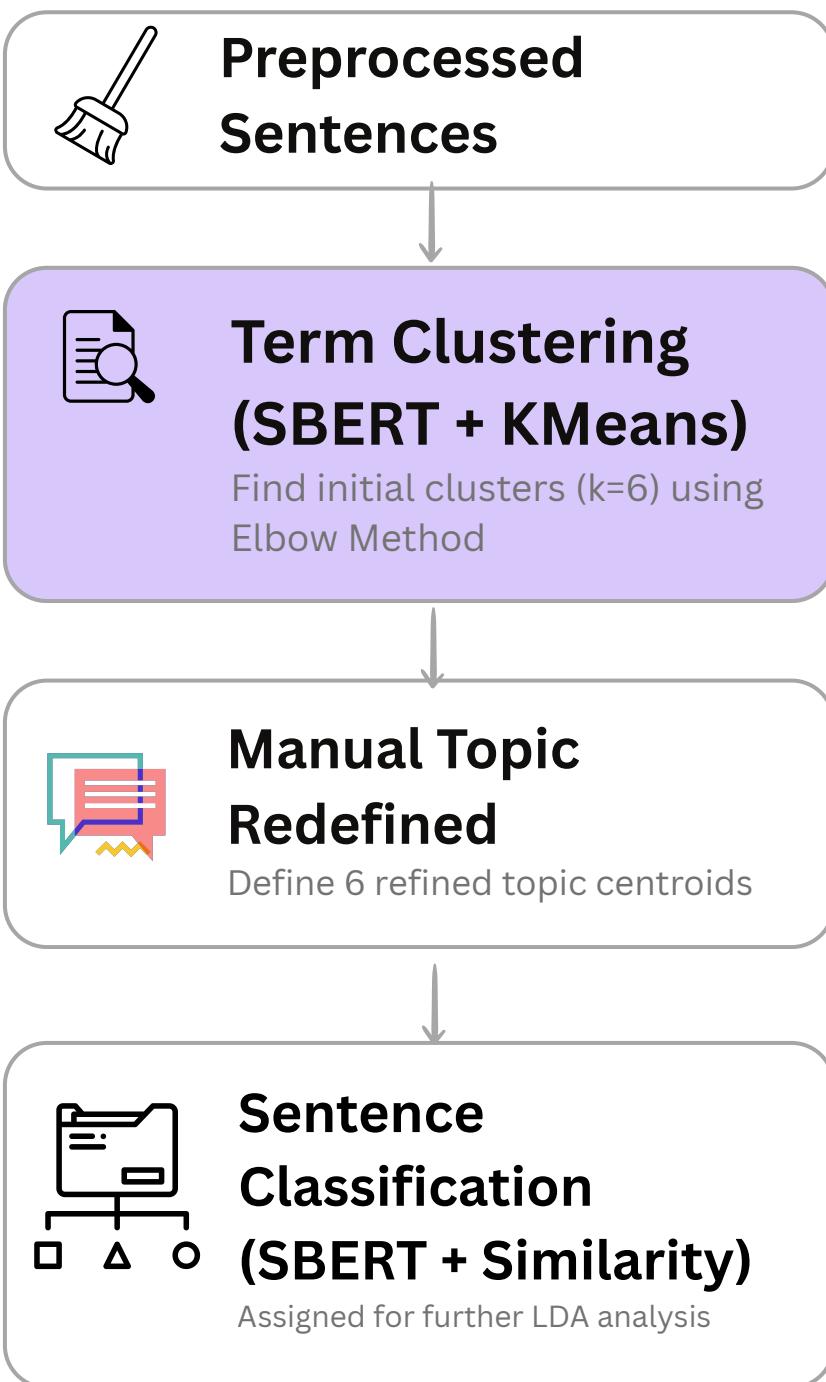
- **Dedicated Q&A Section**

- Create a separate area for questions without requiring a star rating.
- Allows users to seek information without distorting rating data.



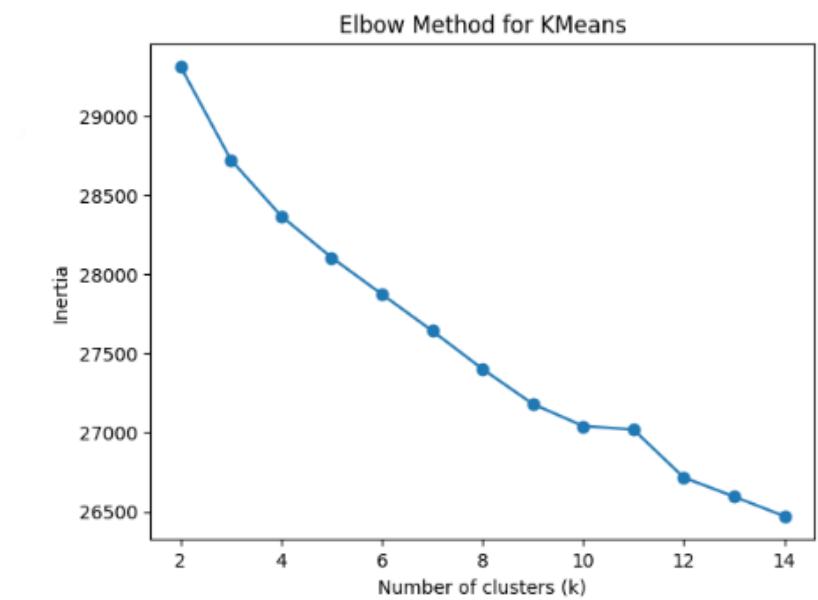
Product Review Global Topics (Stage 1)

Workflow of global topics



Term Clustering: 6 initial clusters

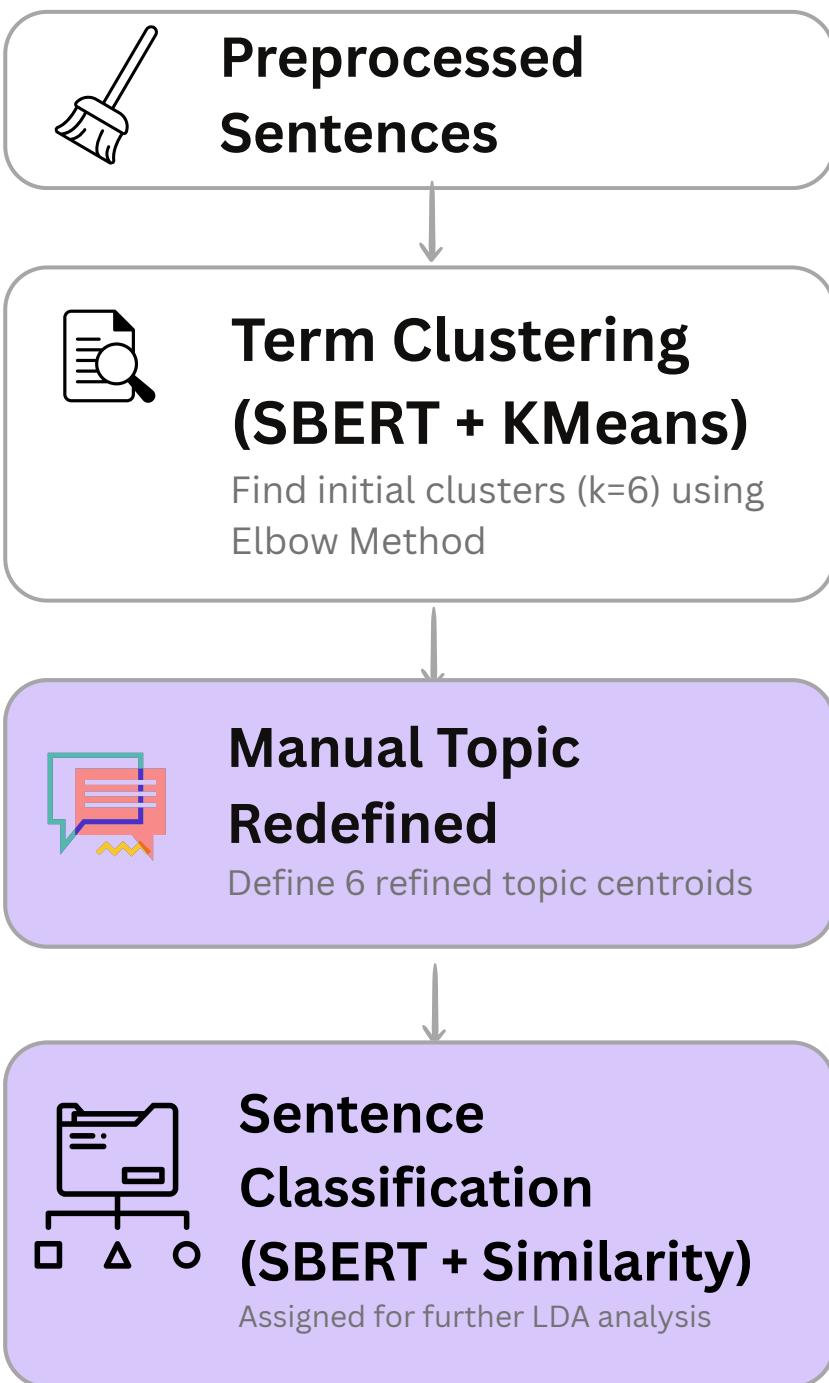
Cluster ID	Representative Name	Top Terms (Synonyms)
1	hydrate	hydrate, cleanse, moisture, hydration, dewy, spray, absorb, cleanse_balm, dry_patch, mist
2	dry	dry, sensitive, smell, clean, soft, price, texture, wash, oil, scent
3	remover	remover, dark_circle, fine_line, ultra_repair, irritate_sensitive, job_remove, large_size, minimize_pore, mini_size, small_size
4	glowy	glowy, dark_spot, huge_fan, rise_petal, high_hope, sit_top, big_fan, extremely_sensitive, stuff_amaze, especially_winter
5	acne	acne, balm, moisturize, acne_prone, peel, moisturizers, redness, exfoliate, retinol, mascara
6	repurchase	repurchase, full_size, price_point, can_not, worth_price, light_weight, worth_money, small_amount, price_tag, waste_money



```
k=2->3: ΔInertia = -590.12
k=3->4: ΔInertia = -354.85
k=4->5: ΔInertia = -261.83
k=5->6: ΔInertia = -232.00
k=6->7: ΔInertia = -231.21
k=7->8: ΔInertia = -241.53
k=8->9: ΔInertia = -220.99
k=9->10: ΔInertia = -139.31
k=10->11: ΔInertia = -22.32
k=11->12: ΔInertia = -303.30
k=12->13: ΔInertia = -120.45
k=13->14: ΔInertia = -123.81
```

Product Review Global Topics (Stage 1)

Workflow of global topics



6 refined global topics based on initial cluster

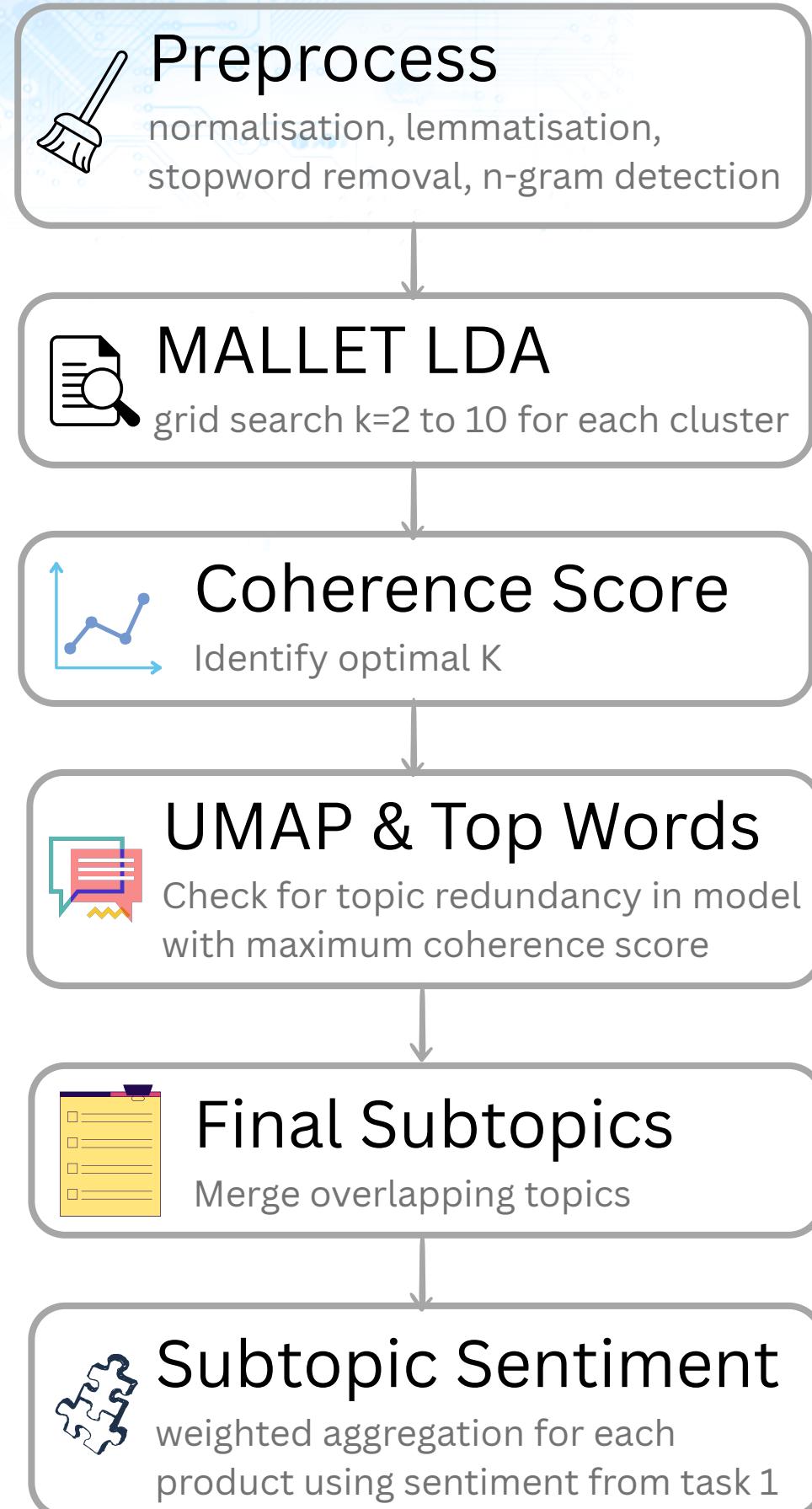
Cluster ID	Cluster Name	Seed Words	Sentence Count
1	dry_hydrate	dry, hydrate, moisture, hydration, absorb, dewy, hydrating	6357 (9%)
2	price_value	price, repurchase, price_point, worth, value, spend, money	1723 (3%)
3	glowy_appearance	glowy, radiant, shine, bright, luminous, lighten, brighten, skin_glow	4594 (7%)
4	smell_scent	smell, scent, fragrance, odor	2352 (3%)
5	acne_effect	acne, break, balm, redness, peel, exfoliate, retinol, breakout, break_out, pimple, blemish, irritation, break_odd	24502 (36%)
6	packaging_size	full_size, mini, sample, packaging, bottle, jar, pump, dispenser, container, travel_size, tube	2037 (3%)

Noted that there are 26209 (39%) of sentence assigned to “unassigned”.

Unassigned: sentence less than 3 words + cosine similarity to cluster centroids < 0.5 + 3 groups that designed to absorb noises (general_effect, usage_routine, purchase_return)

Product Review Subtopics (Stage 2)

LDA & sentiment integration workflow



6 Clusters

dry_hydrate

price_value

glowy_appearance

smell_scent

acne_effect

packaging_size

16 Final Subtopics

1. Hydration & Texture
2. Seasonal & Skin-type Suitability

Perceived Price-value Tradeoff

Glow & Skin Texture Effects

1. Safety & Sensitivity
2. Pleasant Scent
3. Strong/Unpleasant Scent
4. Product Feel & Application

1. Product Routine & Type
2. Skin Concerns & Acne
3. Eye & Makeup Removal
4. Cleansing & Exfoliation
5. Skin Feel & Texture

1. Size, Volume & Travel Value
2. Environmental Impact & Materials
3. Packaging Functionality & Dispensing

Product Review Subtopics (Stage 2)

Example output:



Product	Cluster	Subtopic	FreqWeight (Topic Prominence)	Subtopic Average Sentiment	Subtopic Score (Avg Sentiment * FreqWeight)	Product Average Sentiment
Alpha Beta Extra Strength Daily Peel Pads	dry_hydrate	Hydration & Texture	24.8	0.65	16.2	0.44
Alpha Beta Extra Strength Daily Peel Pads	dry_hydrate	Seasonal & Skin-type Suitability	6.22	0.61	3.8	0.44

- Both subtopics have **high positive sentiment (>0.6)**, **above product average (0.44)** → Aspects in which customers are highly satisfied with (key strengths)
- Hydration & Texture is mentioned far more frequently than Seasonal & Skin-type Suitability, making **hydration effects the core strength**, while **versatility across skin types and seasons** serves as a **supportive strength**.

Product-Level Insights: What Customers Really Think

High Priority Opportunities

HIGH VOLUME + HIGH SENTIMENT (Invest & Amplify)

- **Glow & Skin Texture Effects** - Customer favorite, strong performance
- **Hydration & Texture** - High discussion, positive reception

HIGH VOLUME + LOW SENTIMENT (Urgent Improvement)

- **Skin Concerns & Acne** - 27.5% of conversations, needs formulation work
- **Safety & Sensitivity** - Critical trust issues requiring attention

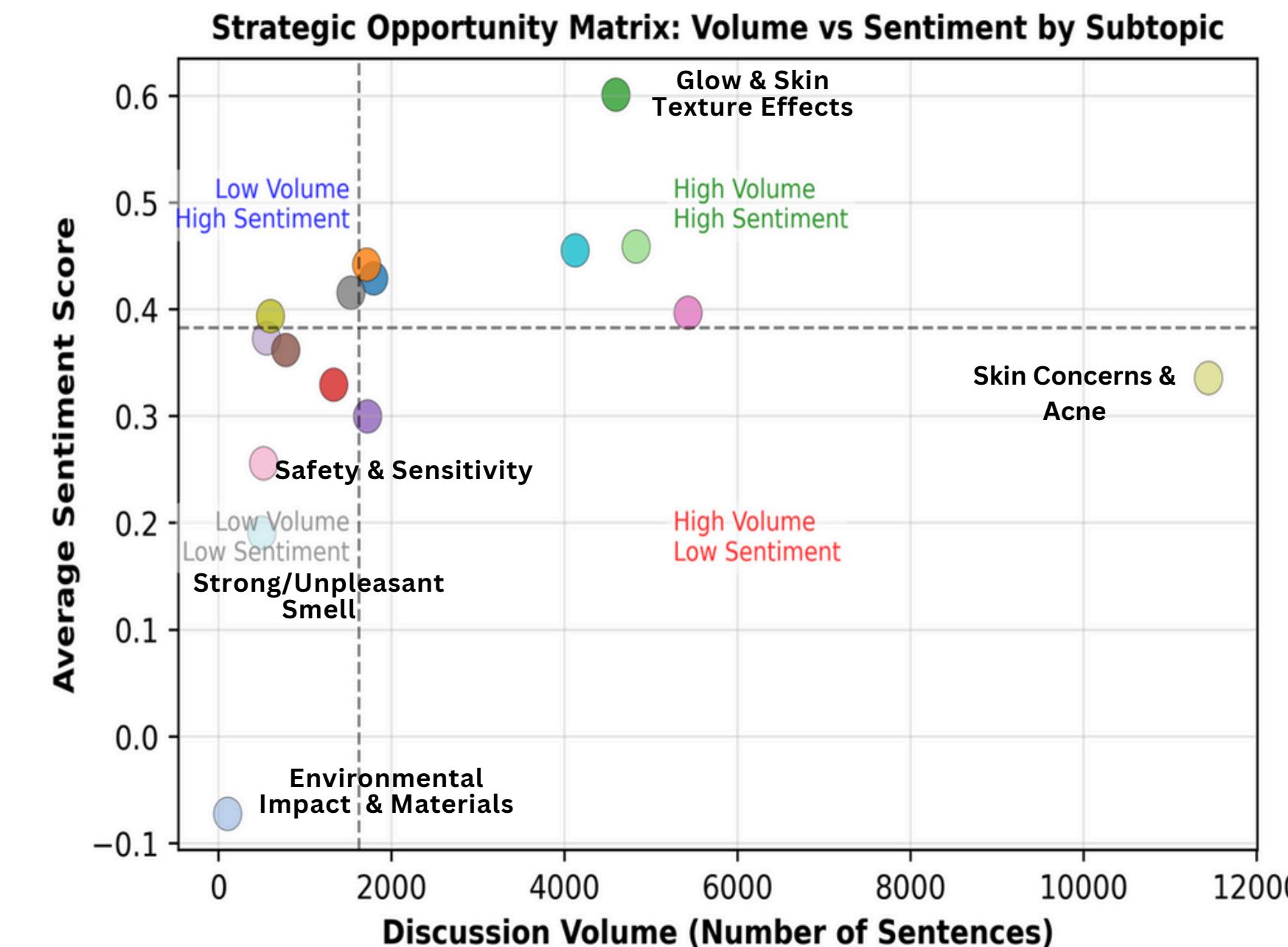
Strategic Focus Areas

LOW VOLUME + HIGH SENTIMENT (Growth Opportunities)

- **Pleasant Scent** - Under-discussed positive attribute
- **Seasonal Suitability** - Niche with positive reception

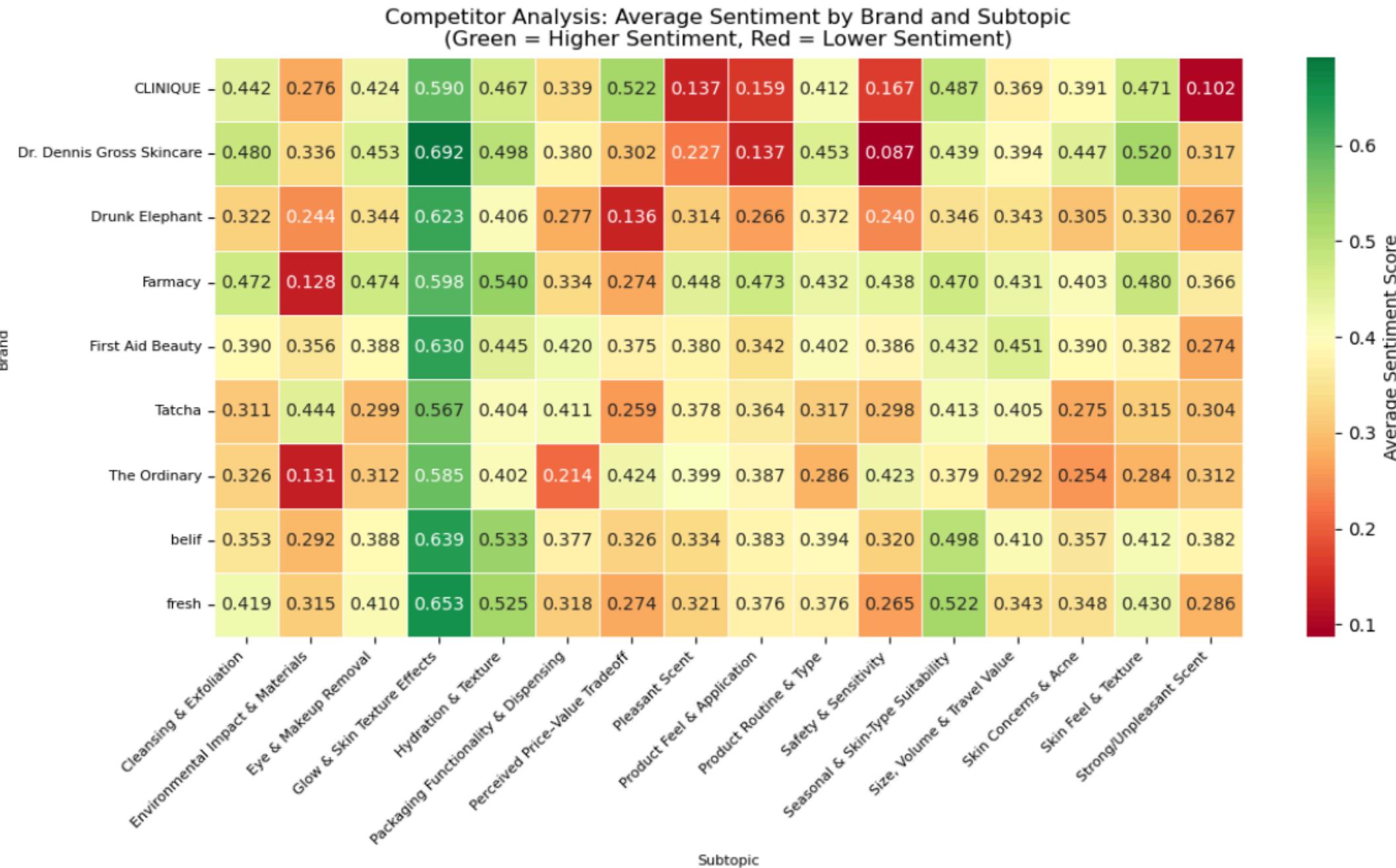
LOW VOLUME + LOW SENTIMENT (Monitor & Optimize)

- **Environmental Impact & Materials** - Emerging concern area
- **Strong/Unpleasant Scent** - Specific but severe pain point



Key Insight: Focus resources on High Volume areas where sentiment improvements will impact the most customers

COMPETITOR ANALYSIS: Where We Win & Lose



AMPLIFY – Leverage Attribute Leaders

- **“Glow & Texture” Power Duo:** Bundle Dr. Dennis Gross (0.692) + Fresh (0.653) in “Pro-Glow” kits.
- **“Hydration” Heroes:** Spotlight Farmacy (0.540) + belif (0.533) as top solutions for dryness.

EXPLOIT – Fix Brand Weaknesses

- **Close the “Sustainability Gap”** (Environmental Impact): Improve perception for Farmacy (0.128) + The Ordinary (0.131).
- **Tackle “Perceived Price-Value” Complaints:** Reframe pricing for Drunk Elephant (0.136) + Tatcha (0.259) via content storytelling.

CAPTURE – Smart Competitive Positioning

- **Value Shoppers:** Suggest CLINIQUE (0.522) as alternative to Tatcha (0.259).
- **Efficacy Shoppers:** Prioritize Dr. Dennis Gross (0.447) over weaker performers in skin concern and acne category.

Review Summarization

Objectives:

Transform fragmented customer reviews into concise summaries. Present product strengths and weaknesses to provide a basis for product decisions.

Hopefully, you find your perfect skin regimen as well.

As far as over the counter retinols go,
is my all-time favorite. I have been using it
this is the best you can get.

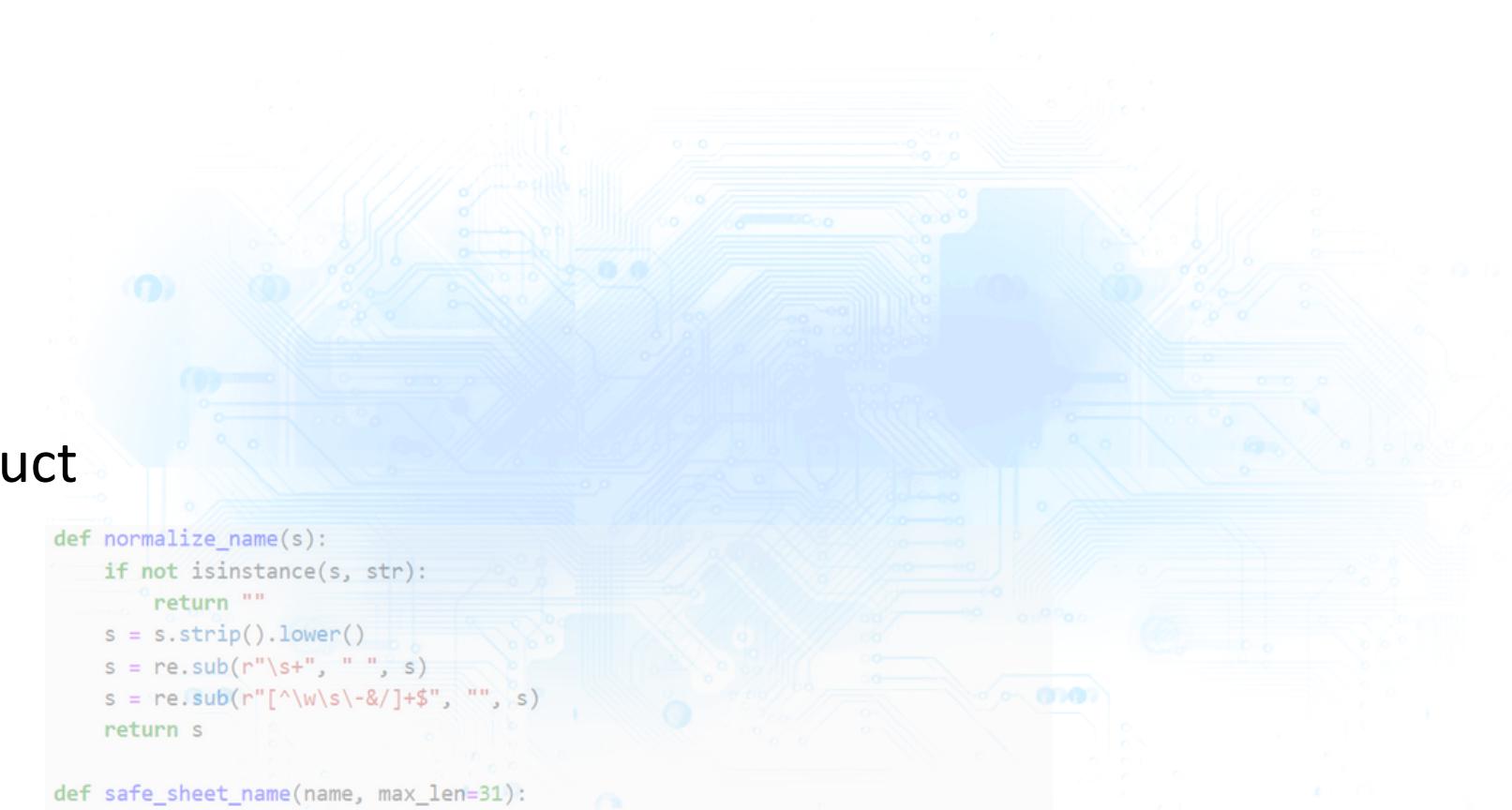
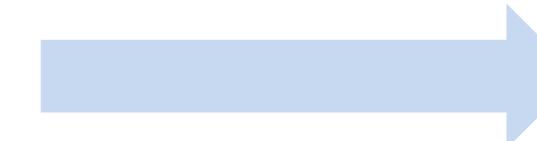
Customer Reviews

literally after using this twice i noticed a difference and after a couple weeks the cystic acne on my chin and consistent texture on my forehead is completely gone.

Best retinol product I have ever used

I have extremely sensitive skin as well, which has made finding a good product a challenge. I'm a skincare junkie, and this is one of the best products I've ever used. This is the BEST product I have ever tried for pore reduction! Highly recommend. The best skincare product I have ever used! This is my absolute favorite. All time favorite brightening cream. After using this for a few weeks, I feel like my skin is so much brighter and more even-toned. Best overnight hydrating cream I've ever used. Highly recommend this product. By FAR the absolute best hydration product I've ever used. This is the best moisturizer I have ever used. I feel like my skin is so much smoother and more hydrated. Love using this on my eyes. It's great for a quick pick-me-up if I'm hungover or really tired. I have dry patchy skin and after I shower I have to remove all the dead skin from my face. This product is the best clean cleanser for dry skin and sensitive skin ! My skin does best with this product. Really one of the best products I have ever used, and one I will repurchase. Makes my skin look the best it ever has. Perfect for summer or anytime you need a quick pick-me-up. This product saves my skin and makes it look the best it has ever been! I love this product. It is by far the best product you will ever use if you have very dry skin and sensitive skin. This was my first time using this cleanser and I so far love it! I love this product. I've tried a couple other cleaning balms and this is the best. super gentle. I use this under the it cosmetics CC cream and it makes my face look so good. This coconut water cream is one of the best moisturizers I have ever used. Love, love, love I use this every night in my skin care. It's so hydrating.

Overall Summary



```
def normalize_name(s):
    if not isinstance(s, str):
        return ""
    s = s.strip().lower()
    s = re.sub(r"\s+", " ", s)
    s = re.sub(r"[\^w\s\-\&\ ]+$", "", s)
    return s

def safe_sheet_name(name, max_len=31):
    return name[:max_len]

def infer_sentiment_column(df):
    """
    Infer the column containing sentiment or compound.
    Preference order:
    - columns containing 'sentiment' or 'compound'
    Returns (colname, scale) where scale is one of {"pm1", "01", "15"} meaning:
    pm1: values roughly in [-1,1]
    01 : values in [0,1]
    15 : values in [1,5] (ratings)
    """
    candidates = []
    for c in df.columns:
        lc = c.lower()
        if "sent" in lc or "compound" in lc:
            if pd.api.types.is_numeric_dtype(df[c]):
                nonnull = df[c].notna().sum()
                if nonnull > 0.5: # at least half filled
                    candidates.append((c, lc))

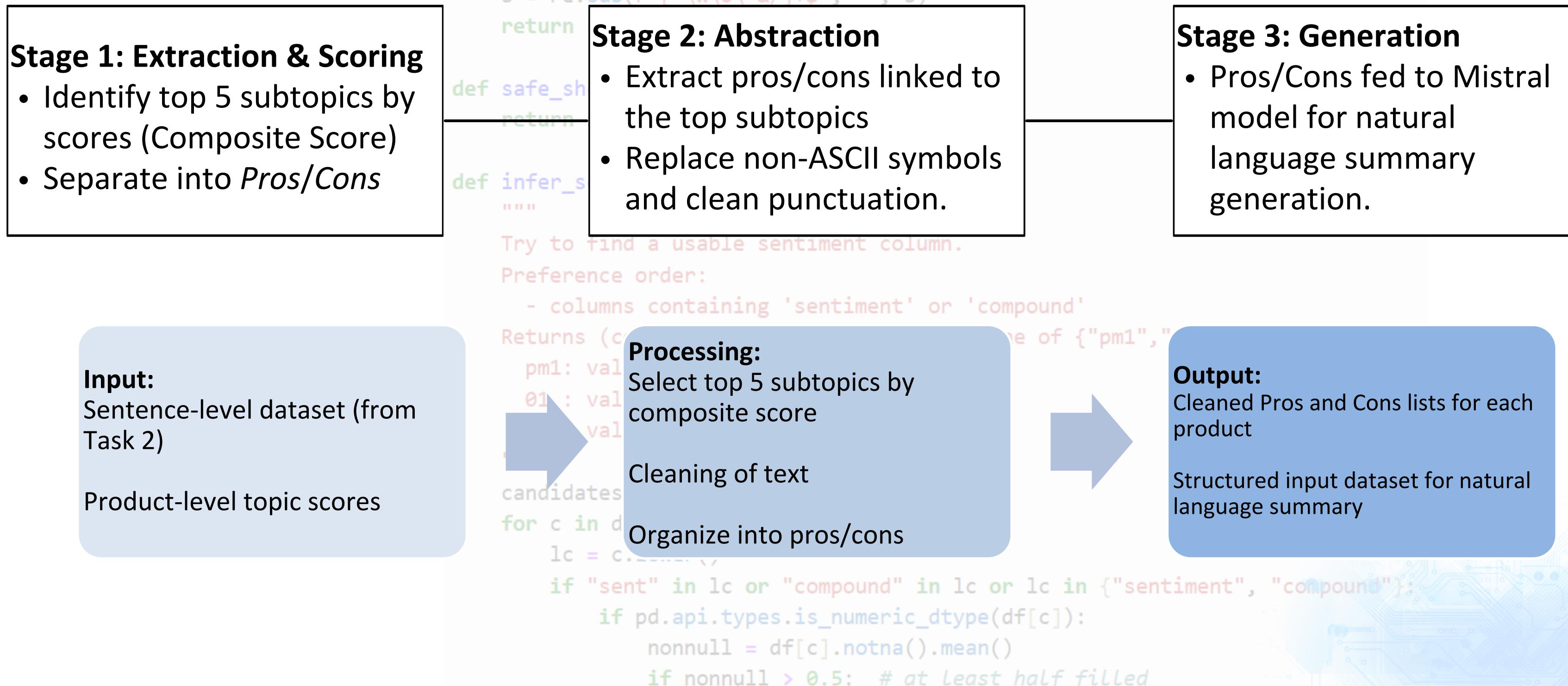
    if len(candidates) == 0:
        raise ValueError("No sentiment or compound columns found in the input DataFrame")
    else:
        return candidates[0]
```

Extraction & Abstraction

Mistral Model Generation

Review Summarization

Hybrid Summarization Pipeline



Review Summarization

Abstraction Outputs

Product	Review	Pros Summary	Cons Summary	Abstractive Summary
A-Passioni Retinol Cream	<p>Pros:</p> <ul style="list-style-type: none">1. "Definitely the most effective skincare product I've ever used!"2. "The best retinol product I've ever used and the only one I've seen results from."3. "I have used other retinols (biossance and dr Dennis, Murad) this is the best I've used.'4. "Out of all my skincare products this one is the one I will not be giving up."5. "The best retinol product I've found." <p>Cons:</p> <ul style="list-style-type: none">1. "It was very disappointing given the high price."2. "Had to stop use which is especially unfortunate given the price tag."3. "way overpriced and made me breakout."4. "I don't mind spending good money on good products but these products are just a waste of money."5. "Not worth it for the price and the strength of the retinol."	Definitely the most effective skincare product I've ever used! The best retinol product I've ever used and the only one I've seen results from. I have used other retinols (biossance and dr Dennis, Murad) this is the best I've used. Out of all my skincare products this one is the one I will not be giving up. The best retinol product I've found.	It was very disappointing given the high price. Had to stop use which is especially unfortunate given the price tag. way overpriced and made me breakout. I don't mind spending good money on good products but these products are just a waste of money. Not worth it for the price and the strength of the retinol.	Overall, users found that many consider it a customer favorite and among the better over-the-counter options. It delivers noticeable improvements to skin texture and clarity. However, a number of customers mentioned next-day breakouts or bumpy skin.

Review Summarization (cont'd)

Generative LLM Model Output

Product	Review	Abstractive Summary	Generated Summary
A-Passioni Retinol Cream	<p>Pros:</p> <ol style="list-style-type: none">1. "Definitely the most effective skincare product I've ever used!"2. "The best retinol product I've ever used and the only one I've seen results from."3. "I have used other retinols (biossance and dr Dennis, Murad) this is the best I've used.'4. "Out of all my skincare products this one is the one I will not be giving up."5. "The best retinol product I've found." <p>Cons:</p> <ol style="list-style-type: none">1. "It was very disappointing given the high price."2. "Had to stop use which is especially unfortunate given the price tag."3. "way overpriced and made me breakout."4. "I don't mind spending good money on good products but these products are just a waste of money."5. "Not worth it for the price and the strength of the retinol."	Overall, users found that many consider it a customer favorite and among the better over-the-counter options. It delivers noticeable improvements to skin texture and clarity. However, a number of customers mentioned next-day breakouts or bumpy skin.	This retinol product has received mixed reviews from customers. Some have reported excellent results, praising it as the most effective skincare product they've ever used and the best retinol product they've tried. They've noticed significant improvements and plan to continue using it. Others, however, have been disappointed with the product, finding it overpriced and not worth the cost. Some users even reported breakouts after using it. Overall, the product has shown to be effective for some, but not for all, and its high price tag has been a point of contention.

Limitations

Task	Limitation	Future Work
Sentiment Analysis	<ul style="list-style-type: none">Struggles with sarcasm, tonal shifts, and mixed emotions, failing to capture nuanced expressions.Unable to link sentiments to specific product attributes, often giving an averaged score that blurs contrast.	<ul style="list-style-type: none">Use ABSA to link sentiment with specific attributes.Incorporate dependency parsing / attention for better context alignment.Build a feedback-driven iterative process to retrain on misclassified samples.
Topic Modeling	<ul style="list-style-type: none">Some semantic overlap and ambiguous subtopics remain despite manual refinement.Current structure (SBERT + LDA) lacks full hierarchy — subtopics not always intuitive.	<ul style="list-style-type: none">Adopt Hybrid Hierarchical Topic Modeling (HTM): Combine the semantic power of SBERT embeddings with structural rigor of HTM to achieve cleaner, nested subtopics and remove ambiguity.
Summary Generation	<ul style="list-style-type: none">Limited dataset coverage as only the top 5 subtopics per product were summarised which may exclude lower-frequency but potentially meaningful feedback.Manual evaluation bias since factual accuracy was assessed manually without automated metrics such as cosine similarity	<ul style="list-style-type: none">Metric-based Evaluation: Incorporate automatic coherence and factual alignment metrics.Adaptive Summarisation: Enable user-specific summary tuning.Broader Coverage: Expand to include more subtopics and lower-frequency reviews to improve representativeness.

Project Extensions



Achievement

- Built a **scalable NLP pipeline** turning Sephora's unstructured reviews into **actionable business intelligence**



Future Extensions

- **Temporal Analysis:** Track the evolution of subtopic prevalence and sentiment over time to uncover seasonal trends and recurring patterns, informing inventory and campaign timing.
- **Link to Business KPIs (ROI Quantification):** Use causal models to quantify **financial impact of addressing specific customer concerns** (e.g., how improving "Safety & Sensitivity" sentiment affects sales, conversion, and retention).
- **Customer Segmentation:** Apply clustering to aggregated subtopic sentiment and prominence scores to **group customers** based on their **product concerns and priorities** (e.g., "Safety-First Buyers," "Eco-Conscious Shoppers"). This enables **personalized marketing and product recommendations**.

Thank You

