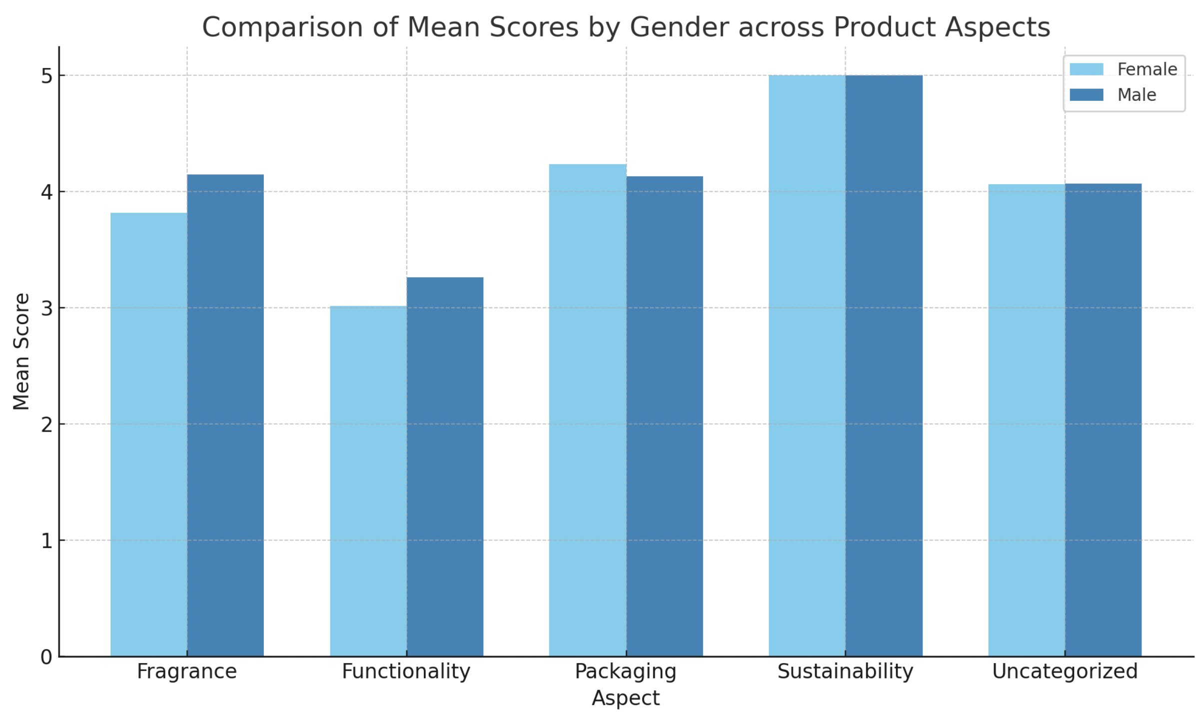
Section 14: Aspect-Based Sentiment Analysis

1. **Aspect-Sentiment Dataset**
    - Columns: Respondent_ID, Question_ID, Aspect, Raw_Text, Sentiment_Score

https://docs.google.com/spreadsheets/d/1ejnuioXFQa2MBZ4P95fUspL6CNPP3AqmDqG_eoeqtSc/edit?gid=469878737#gid=469878737

2. **Aggregated Sentiment Table**
    - Rows: Aspect × Cohort (Packaging × Gender), Columns: Mean_Sentiment, n



Insights from the Chart: Gender Comparison by Product Aspect Ratings

1. **Fragrance**:
    - Males rated fragrance slightly higher (4.15) than females (3.82).
    - Suggests fragrance may play a marginally stronger role in male preference.
2. **Functionality**:

- Both genders rated it lower than other aspects.
- Males rated it slightly higher (3.26) than females (3.02), indicating minor gender-based difference in perceived effectiveness.

3. **Packaging**:
   - Females rated packaging higher (4.24) than males (4.13), suggesting more visual or usability sensitivity in packaging appeal among females.
4. **Sustainability**:
   - Perfect score (5.0) from both genders.
   - Indicates universal appreciation and alignment on sustainability as a valued brand aspect.
5. **Uncategorized**:
   - Similar scores (Female: 4.06, Male: 4.07).
   - No notable difference, implying consistent impressions across non-classified brand features.

Aspect	A1	A2	B1	B2	C1	C2
Fragrance	3.50	4.40	3.67	3.97	4.11	4.13
Functionality	3.18	2.71	3.43	2.98	3.08	3.17
Packaging	4.04	4.40	4.13	4.17	4.23	4.19
Sustainability	5.00	5.00	5.00	5.00	5.00	5.00

Uncategorized	4.04	4.00	4.07	4.01	4.16	4.13

Insights:

1. **Sustainability**:
   - All NCCS segments rated this a **perfect 5.0**, indicating strong universal appeal of sustainability features across all socioeconomic classes.
2. **Fragrance**:
   - Highest rating from **A2 (4.40)** and relatively lower in **A1 (3.50)** and **B1 (3.67)**.
   - Suggests that fragrance appeal resonates more with **upper-middle and lower NCCS segments**.
3. **Functionality**:
   - **B1 (3.43)** leads slightly, while **A2 (2.71)** is the lowest.
   - Functional benefits are better appreciated among **middle NCCS groups**, but generally rated modestly across the board.
4. **Packaging**:
   - **A2 (4.40)** and **C1 (4.23)** rated highest.
   - Indicates packaging appeal is particularly strong for **visually-inclined or value-conscious consumers**.
5. **Uncategorized Aspects**:
   - Fairly consistent across segments (4.00–4.16 range).
   - No strong variance, indicating general uniformity in how non-specific brand features are perceived.

none of the aspect groups had enough data across NCCS segments to perform ANOVA or Kruskal–Wallis. This can happen if

Description:

Section 14: Aspect-Based Sentiment Analysis (ABSA)

Scope: Identify sentiment tied to specific product aspects in open-ended responses using a BERT-based ABSA pipeline.

Objectives

- Quantify sentiment for aspects like *Packaging*, *Fragrance*, *Ingredients*, *Sustainability*, etc.
- Tag each response with relevant aspect(s).
- Use a BERT model to predict sentiment for each (aspect, response) pair.
- Analyze sentiment by demographic groups: Age, Gender, NCCS segment.

Analysis Tasks

Task	Details	Method
1. Aspect Dictionary	- Define key product aspects: Packaging, Fragrance, Ingredients, Functionality, Sustainability. - Create keyword sets for each aspect.	Manual + lookup table

2. Aspect Tagging	- Use regex or keyword match to tag each response with one or more aspects.	`python import pandas as pd import re aspects = { "Packaging": ["packaging", "bottle", "design", "label"], "Fragrance": ["smell", "fragrance", "scent"], "Ingredients": ["ingredient", "content", "chemical", "natural"], "Functionality": ["effect", "result", "clean", "lather"], "Sustainability": ["eco", "green", "recyclable", "sustainable"] } def tag_aspects(text): tagged = [] for aspect, keywords in aspects.items(): if any(re.search(rf'\b{k}\b', text.lower()) for k in keywords): tagged.append(aspect) return tagged df = pd.read_csv("openends_responses.csv") df["Aspects"] = df["Raw_Text"].apply(tag_aspects)`
3. Sentiment Scoring	- Use pretrained model: `nlptown/bert-base-multilingual-uncased-sentiment`. - Predict sentiment score (1–5) for each (aspect, snippet).	`python from transformers import AutoTokenizer, AutoModelForSequenceClassification from transformers import pipeline tokenizer = AutoTokenizer.from_pretrained("nlptown/bert-base-multilingual-uncased-sentiment") model = AutoModelForSequenceClassification.from_pretrained("nlptown/bert-base-multilingual-uncased-sentiment") sentiment_pipeline = pipeline("sentiment-analysis", model=model, tokenizer=tokenizer) def score_sentiment(text): result = sentiment_pipeline(text[:512])[0] return int(result['label'].split()[0]) expanded = df.explode('Aspects') expanded['Sentiment_Score'] = expanded['Raw_Text'].apply(score_sentiment)`

4. Aggregation	- Compute average sentiment per aspect per respondent and per demographic cohort.	`python # Mean sentiment by aspect and gender grouped = expanded.groupby(['Aspects', 'Gender'])['Sentiment_Score'].agg(['mean', 'count']).reset_index() grouped.columns = ['Aspect', 'Gender', 'Mean_Sentiment', 'n']`
5. Statistical Testing	- Use ANOVA or Kruskal–Wallis to test sentiment variation across Age or Gender.	`python from scipy.stats import f_oneway age_groups = expanded.groupby(['Age_Bucket', 'Aspects'])['Sentiment_Score'].apply(list).unstack() for aspect in age_groups.columns: scores = [group for group in age_groups[aspect].dropna()] stat, p = f_oneway(*scores) print(f"{aspect} — ANOVA p-value:", p)`
6. Visualization	- Heatmap of sentiment scores (Aspect × Demographic) - Bar charts for top sentiment differences	`python import seaborn as sns import matplotlib.pyplot as plt heat_data = expanded.groupby(['Aspects', 'Gender'])['Sentiment_Score'].mean().unstack() sns.heatmap(heat_data, annot=True, cmap="RdYlGn", center=3) plt.title("Mean Sentiment Score by Aspect and Gender") plt.show() sns.barplot(data=grouped, x='Aspect', y='Mean_Sentiment', hue='Gender') plt.xticks(rotation=45) plt.title("Sentiment by Aspect and Gender") plt.tight_layout() plt.show()`

Deliverables

Aspect-Sentiment Dataset

File: `aspect_sentiment_data.csv`
Columns: `Respondent_ID`, `Question_ID`, `Aspect`, `Raw_Text`, `Sentiment_Score`

Aggregated Sentiment Table

Rows: Each combination of Aspect × Demographic Group
Columns: `Mean_Sentiment`, `n`
Example:

Aspect	Gender	Mean_Sentiment	n
Packaging	Male	3.8	124
Fragrance	Female	4.1	138

Visuals

- **Heatmap:** Mean sentiment across Aspects and Demographics
- **Bar Charts:** Aspect sentiments by Gender, Age, NCCS

ABSA Notebook

- Fully documented with code for:
  - Aspect dictionary creation
  - Tagging logic
  - Model-based sentiment prediction
  - Group-level aggregation
  - Statistical tests and visualizations

Summary Report

- **Top Positive Aspects:** Sorted by average sentiment score
- **Top Negative Aspects:** Flag aspects with low average sentiment
- **Group-Level Insights:** Key differences by Age, Gender, NCCS
- **Recommendations:** e.g., “Improve packaging design for 18–25 segment”