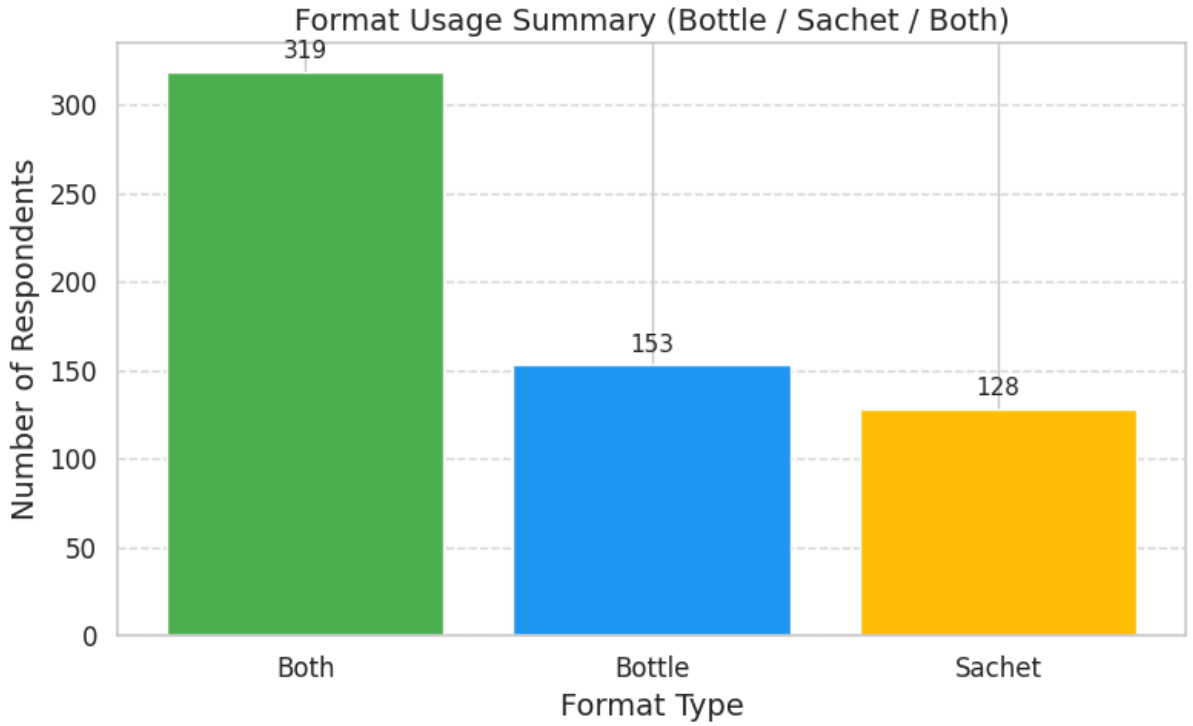
Packaging Format Preferences (Q39–Q41)

1. **Format Usage Summary**



Format Usage Summary (Bottle / Sachet / Both)

- **Both**: 319 respondents (53.2%)
- **Bottle only**: 153 respondents (25.5%)
- **Sachet only**: 128 respondents (21.3%)

🔍 Key Insights:

1. **Majority Use Both Formats**
   Over half the consumers (53.2%) reported using **both sachets and bottles**, indicating flexibility in product usage. This suggests users adapt based on convenience, price, or availability.

2. **Bottles Slightly More Preferred Than Sachets**
   Among users who prefer a single format:

- **Bottles (25.5%)** are used more than **sachets (21.3%)**.
- Bottles likely appeal to regular users who prioritize long-term value or aesthetics.
- Sachets may appeal to price-sensitive or trial-based buyers.

3. **Implication for Brands**
   - **Maintain dual-format availability**: Brands should continue offering both formats as a significant portion uses both.
   - **Tailor messaging by format**:
     - Bottles: Emphasize premium, value-for-money, and sustainability.
     - Sachets: Highlight affordability, portability, and convenience.
4. **Marketing Strategy Opportunity**
   - For brands trying to **upgrade sachet users to bottles**, emphasize long-term savings or improved formulation.
   - For bottle users, push **value packs** or bundled offers.

2. **Reason Themes**

Reason Themes – Format Preference (Q40 & Q41)

LDA Topic 1 (Bottle Preference)

Top Words: bottle, want, available, avoid, buy

Interpretation & Insights:

- Respondents preferring **bottles** emphasized:
  - **Desire for availability**: Bottles are perceived as reliable and easier to find in stores.
  - **Buying behavior**: Suggests a more **planned purchase** approach — bottles are often bought during regular shopping trips.

    - **Avoidance reasoning**: Users may **avoid sachets** due to waste, messiness, or smaller quantity.
    - **Value orientation**: Bottles may be linked to **bulk buying** and **long-term value**.

Implication: Users choosing bottles are likely committed, value-driven buyers looking for convenience and consistent product use.

LDA Topic 2 (Sachet Preference)

Top Words: sachets, like, use, carry, family

Interpretation & Insights:

- Respondents who prefer **sachets** associate them with:
    - **Ease of use and portability**: Sachets are convenient to carry while traveling or for shared use within the family.
    - **Affordability and trial**: Sachets may appeal to price-sensitive segments or those who like trying different variants.
    - **Household dynamics**: Mention of *family* implies shared use, possibly in multi-member or lower NCCS households.

Implication: Sachet users represent a cost-conscious, convenience-seeking segment that values mobility and flexibility in usage.

3. **Regression Output**
    - Multinomial logit results with RRRs, 95% CIs, and significance

Optimization terminated successfully.

Current function value: 0.677235

Iterations 4

✅ Multinomial Logistic Regression Report:

	RRR	p-value
const	1.129594	0.740163
Age_23 to 29 years old	0.939398	0.796292
Age_30 to 39 years old	1.279703	0.308417
Age_40 to 45 years old	1.625690	0.067177
Gender_Male	1.148680	0.406835
NCCS_A2	0.431340	0.027733
NCCS_B1	0.450875	0.013165
NCCS_B2	0.471021	0.019027
NCCS_C1	0.610792	0.156858
NCCS_C2	0.412111	0.011371
Usage_occasionally	1.280463	0.230029
Usage_weekly	1.260490	0.257715
Dandruff_Moderate	1.282699	0.222486
Dandruff_Severe	1.185401	0.413975

Overview:

This model predicts the **likelihood of preferring the test product** over the control product based on demographic and behavioral variables. The key metric here is the **RRR (Relative Risk Ratio)**:

- **RRR > 1** means **higher likelihood** of choosing the test product.
- **RRR < 1** means **lower likelihood**.
- A **p-value < 0.05** indicates statistical significance.

Key Significant Predictors:

Variable	RRR	p-value	Insight
NCCS_A2	0.431	0.028	Statistically significant. Respondents in NCCS A2 are **less likely** to prefer the test product (vs. the control).
NCCS_B1	0.451	0.013	Significant negative impact. NCCS B1 respondents are **less likely** to choose the test.
NCCS_B2	0.471	0.019	Significantly less likely to prefer the test product. Indicates a consistent trend among mid-income groups.
NCCS_C2	0.412	0.011	Strong negative impact on test preference. Suggests **low-income segments resist switching** to the test product.

Interpretation:

- Across **all NCCS segments below A1**, there is a statistically significant **lower preference for the test product**. This suggests that the **test product may not resonate with price-sensitive or value-driven segments**.

Non-Significant Predictors (but worth noting):

- **Age 40–45 (RRR = 1.63, p = 0.067)**: Approaches significance. Older respondents may be **more inclined** to try the test product but not conclusively.
- **Usage Frequency, Gender, Dandruff Severity**: No significant effect, suggesting that **current usage patterns or scalp condition are not strongly driving preference**.

Strategic Implications:

- **Affluent consumers (likely NCCS A1)** show **relatively higher affinity** for the test product, possibly due to perceived quality or innovation.

- **Middle and lower-income groups** are hesitant—potentially due to **price sensitivity**, **perceived value**, or **brand trust**.

- Future positioning of the test product should:

    - Emphasize **value-for-money or performance benefits** for mass-market segments.

    - Consider **differentiated packaging or trial formats** to attract lower NCCS groups.

    - Investigate qualitative drivers behind NCCS resistance.

4. **“Both” Segment Profile**

    - Demographic and usage cadence summary chart (e.g., radar or bar)

    - Key motivational themes distribution

Short Insights from Motivational Themes (Text Clusters):

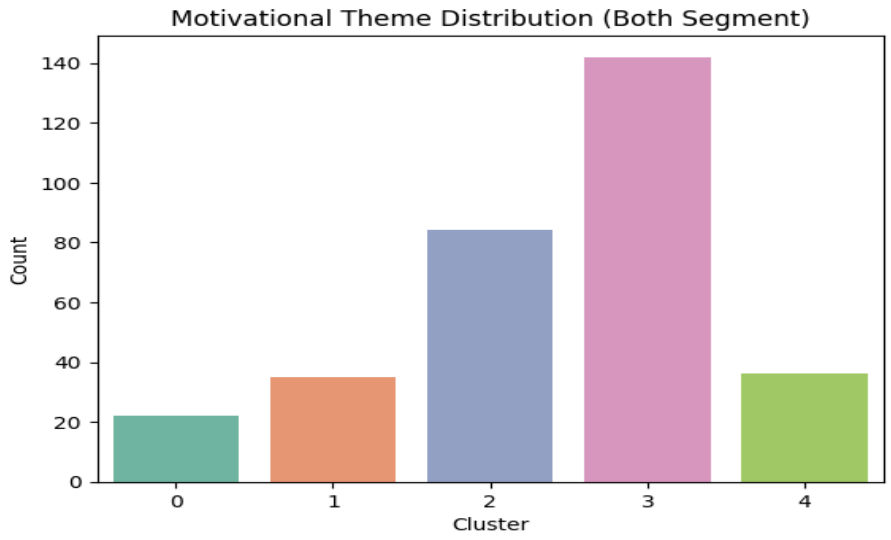
1. **Cluster 0 – Trust & Efficacy:**
   Consumers value **effective dandruff control** and trust in **well-known, long-used brands**.

2. **Cluster 1 – Scalp Care:**
   Motivated by **cleanliness and irritation-free experience**, indicating preference for **mild yet effective products**.

3. **Cluster 2 – Hair Suitability & Strength:**
   Focus on **personal fit**, strengthening roots, and **shiny hair** – signals a **beauty-driven segment**.

4. **Cluster 3 – Style & Convenience:**
   Prioritize **volume, texture**, and ease of use; **styling and packaging** play a key role here.

5. **Cluster 4 – Brand Ethics & Reputation:**
   Drawn to **trusted, eco-conscious, and cruelty-free brands**, showing rising **ethical consumerism**.



Description:

Chapter 8B: Packaging Format Preferences (Q39–Q41)

🎯 Objective

To understand consumer preferences across shampoo packaging formats—bottle, sachet, or both—by analyzing choice distribution, usage rationales, and modeling demographic and behavioral influences. The study also maps associations between format and rationale using correspondence analysis, and profiles dual-format (“both”) users for targeting.

8B.1 Format Usage Summary

📌 Objective:

Quantify frequency and share of format choices.

🛠️ Code:

python
CopyEdit

```
import pandas as pd
```

```
import seaborn as sns
import matplotlib.pyplot as plt

# Q39: 'format' column contains Bottle, Sachet, or Both
format_counts = df['format'].value_counts().reset_index()
format_counts.columns = ['Format', 'Count']
format_counts['Percentage'] = (format_counts['Count'] /
format_counts['Count'].sum()) * 100

# Display table
print(format_counts)

# Bar Chart
sns.barplot(x='Format', y='Percentage', data=format_counts)
plt.title("Shampoo Packaging Format Preference")
plt.ylabel("Percentage (%)")
plt.show()
```

8B.2 Theming Open-Ended Reasons (Q40 & Q41)

📌 Objective:

Extract top 5 themes for sachet use among bottle and mixed-format users.

🛠️ Code (Text Preprocessing & Topic Modeling):

python
CopyEdit

```
from bertopic import BERTopic
from sklearn.feature_extraction.text import CountVectorizer

# Combine open-ends
q40_texts = df['Q40'].dropna().astype(str).tolist()
q41_texts = df['Q41'].dropna().astype(str).tolist()

# Apply BERTopic
topic_model_q40 = BERTopic()
topics_q40, _ = topic_model_q40.fit_transform(q40_texts)

topic_model_q41 = BERTopic()
topics_q41, _ = topic_model_q41.fit_transform(q41_texts)
```

```
# Get top 5 topics
print(topic_model_q40.get_topic_info().head(5))
print(topic_model_q41.get_topic_info().head(5))

# Optional: visualize topics
topic_model_q40.visualize_barchart(top_n_topics=5)
topic_model_q41.visualize_barchart(top_n_topics=5)
```

8B.3 Multinomial Logistic Regression

📌 Objective:

Model format choice using demographics and behavior predictors.

🛠️ Code:

python
CopyEdit

```
import statsmodels.api as sm
import pandas as pd
from sklearn.preprocessing import LabelEncoder

# Prepare data
df_model = df[['format', 'age_group', 'gender', 'nccs',
'usage_freq', 'hair_severity']].dropna()

# Encode categorical predictors
le_format = LabelEncoder()
df_model['format_encoded'] =
le_format.fit_transform(df_model['format'])

# Define predictors and encode dummies
X = pd.get_dummies(df_model[['age_group', 'gender', 'nccs',
'usage_freq', 'hair_severity']], drop_first=True)
y = df_model['format_encoded']

# Fit model
model = sm.MNLogit(y, sm.add_constant(X)).fit()
print(model.summary())
```

```
# Extract Relative Risk Ratios (RRR)
rrr = model.params.apply(lambda x: x.map(lambda y: round(np.exp(y),
2)))
print("Relative Risk Ratios:\n", rrr)
```

8B.4 Correspondence Analysis

📌 Objective:

Visualize associations between packaging format and reason themes.

🛠️ Code:

python
CopyEdit

```
from prince import CA

# Create contingency table from theme presence by format
contingency = pd.crosstab(df['format'], df['reason_theme'])  #
Assume 'reason_theme' is pre-coded
ca = CA(n_components=2, n_iter=10, copy=True, check_input=True,
engine='auto')
ca = ca.fit(contingency)

# Row and column coordinates
row_coords = ca.row_coordinates(contingency)
col_coords = ca.column_coordinates(contingency)

# Heatmap plot
sns.heatmap(ca.transform(contingency), cmap='coolwarm')
plt.title("Correspondence Analysis Heatmap: Format × Reason")
plt.show()

# Optional biplot
ca.plot_coordinates(X=contingency, show_row_labels=True,
show_column_labels=True)
```

8B.5 Profiling “Both” Format Users

📌 Objective:

Compare “Both” users with others on demographics and motivations.

🛠️ Code:

python
CopyEdit

```
# Filter “Both” users
both_users = df[df['format'] == 'Both']

# Demographics summary
demo_summary = both_users[['age_group', 'gender',
'nccs']].value_counts(normalize=True).unstack()
print("Demographics of 'Both' Users:\n", demo_summary)

# Purchase cadence and motivations
cadence_summary = both_users[['purchase_freq',
'purchase_store_type']].value_counts(normalize=True)
print("Purchase Cadence Summary:\n", cadence_summary)

# Compare with bottle- and sachet-only
grouped = df.groupby('format')[['purchase_freq',
'purchase_store_type']].value_counts(normalize=True)
print("Comparative Summary:\n", grouped)

# Radar chart prep (optional)
# Requires conversion to format suitable for radar (not shown here
but can be done)
```
