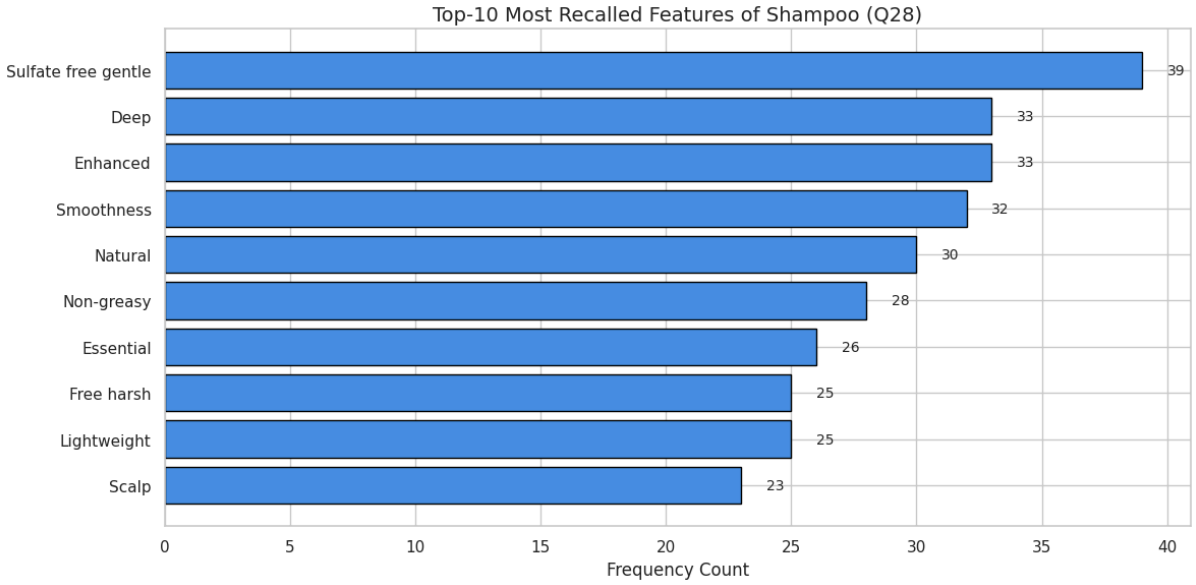
SECTION 7: Product Evaluation (Q28–Q31)

Scope: Leverage Q28–Q31 to understand which product features respondents recall, how they rate the brand, and their intent to retry based on packaging change. Link these attitudes to overall satisfaction (Q30) and future behavior (Q36).

1. **Text Analysis Summary**

Table of top-10 features from Q28/Q29 with frequency counts

Q28 What do you remember most about this shampoo?



The bar chart represents the **top-10 most frequently recalled features** of the shampoo by respondents when asked what stood out to them most:

- **"Sulfate free gentle"** is the most cited attribute (**39 mentions**), indicating strong consumer sensitivity toward mild and chemical-free formulations.

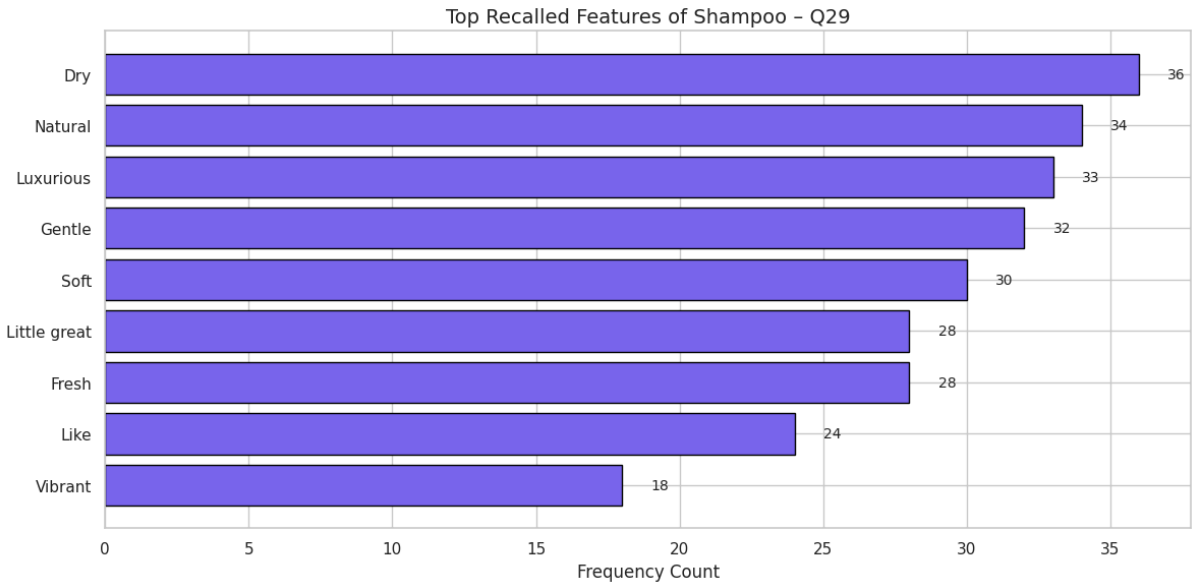
- Features like **"deep"** and **"enhanced"** (each with 33 mentions) suggest recall of performance-related claims such as deep nourishment or improvement in hair texture.

- **"Smoothness"** (32), **"Natural"** (30), and **"Non-greasy"** (28) reflect favorable tactile and sensory experiences.

- Lower down, **"Essential"**, **"Free harsh"**, and **"Lightweight"** (25–26 mentions) point to clean, effective, and user-friendly packaging language.

- **"Scalp"** (23) also appeared often, implying awareness of scalp-targeted benefits.

Q29 What do you remember most about this shampoo?



This visualization displays the **top recalled impressions** about the shampoo based on Q29 responses:

- **"Dry"** was the most mentioned feature (**36 mentions**), which may reflect either a desirable dryness (non-greasy feel) or a negative perception—this should be clarified in qualitative follow-ups.

- **"Natural"** (34) and **"Luxurious"** (33) highlight strong emotional and product experience cues, suggesting aspirational and wellness-linked branding.

- **"Gentle"** (32) and **"Soft"** (30) emphasize mildness and after-use effects on hair texture.

- **"Little great"** (28) and **"Fresh"** (28) may indicate that even small amounts of the product yield satisfying freshness—indicating product efficiency.

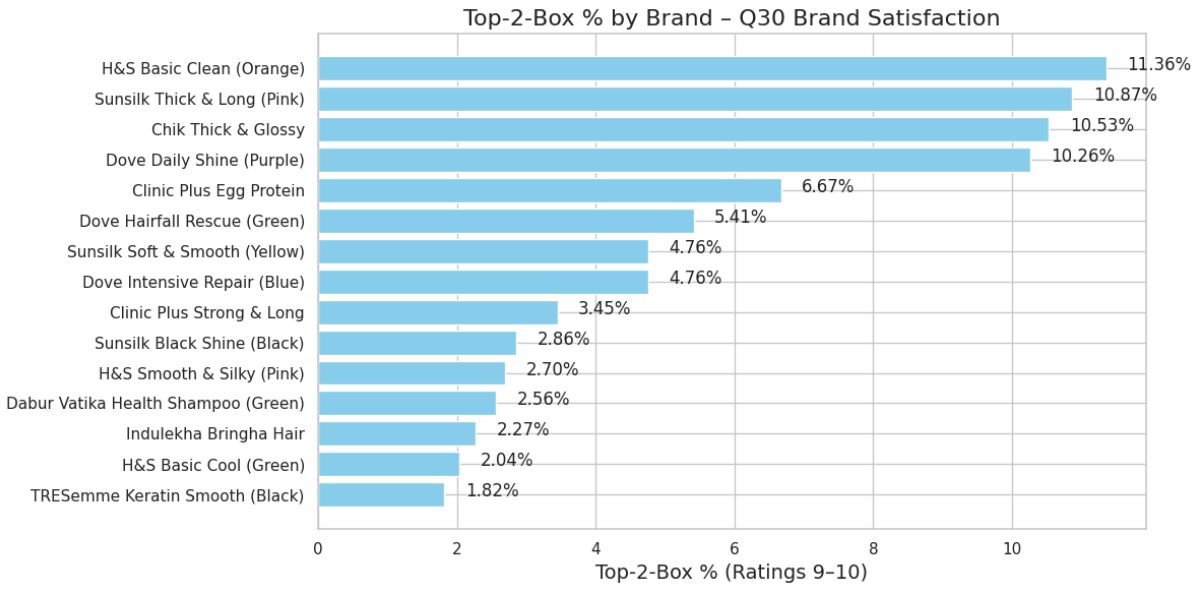
- **"Like"** (24) implies an overall positive recall though it's abstract—useful for sentiment analysis clustering.

- **"Vibrant"** (18) touches on visual or emotional vitality associated with the product, possibly linking to packaging or post-use shine.

2. **Satisfaction Metrics**

% Top-2-Box on **Q30 How would you rate this brand overall on a scale of 1 to 10?**

Top-2-Box (ratings 9–10) for Q30 = 5.50% (33 out of 600)



Summary Insights:

- The overall **Top-2-Box satisfaction is low (5.5%)**, indicating that only a small segment rated their chosen brand as excellent (9 or 10).

- **H&S Basic Clean (Orange)** leads with the highest satisfaction at **11.36%**, followed by **Sunsilk Thick & Long (Pink)** at **10.87%**, and **Chik Thick & Glossy** at **10.53%**.

- Most premium or functional brands (e.g., **Indulekha**, **TRESemme**, **Dabur Vatika**, **H&S Cool**, **Sunsilk Black**) showed **below-average satisfaction** (<3%), which may suggest unmet expectations.

- Dove variants show **moderate satisfaction**, with **Daily Shine (10.26%)** outperforming **Hairfall Rescue** and **Intensive Repair**.

- **Clinic Plus Egg Protein** outperformed its **Strong & Long** counterpart (6.67% vs. 3.45%).

Strategic Implications:

- **Product development teams** should investigate why satisfaction is low across many variants, especially among premium brands.
- **High-performing variants** (like H&S Basic Clean and Sunsilk Thick & Long) can be leveraged as benchmarks for messaging and formulation.
- **Marketing should realign expectations** vs. perceived performance for low-satisfaction brands to reduce disappointment gaps.
- Further **qualitative insights** may reveal what drives “delight” (i.e., 9–10 ratings) for the top-performing variants.

Top-2-Box % by Scalp Segment:

Scalp Condition	Top-2-Box %
Mild	5.10%
Moderate	6.76%
Severe	4.57%

3. **Model Outputs**
    - Ordinal logistic regression results (coefficients, p-values)
    - Correlation summary (ρ or r, p)

OrderedModel Results

==========

Dep. Variable: Q30 Log-Likelihood: -1105.0

Model: OrderedModel AIC: 2228.

Method: Maximum Likelihood BIC: 2267.

Date: Tue, 20 May 2025

Time: 08:41:17

No. Observations: 600

Df Residuals: 591

Df Model: 1

==========

	coef	std err	z	P>|z|	[0.025	0.975]
Q27	0.0612	0.063	0.973	0.331	-0.062	0.185
2/3	-4.3435	0.485	-8.964	0.000	-5.293	-3.394
3/4	0.3485	0.252	1.382	0.167	-0.146	0.843
4/5	0.4425	0.119	3.711	0.000	0.209	0.676
5/6	0.0324	0.088	0.370	0.711	-0.139	0.204
6/7	0.0973	0.071	1.373	0.170	-0.042	0.236
7/8	0.0948	0.084	1.134	0.257	-0.069	0.259
8/9	0.2115	0.123	1.726	0.084	-0.029	0.452
9/10	0.3790	0.214	1.774	0.076	-0.040	0.798

==========

Insights from Ordinal Logistic Regression:

1. **Q27 (Packaging-driven Reuse Intention) is not a significant predictor** of Q30 (Overall Rating).
    - Coefficient: 0.0612, p = 0.331
    - This suggests that even if users say they would continue to use the shampoo if only packaging changed, it **doesn't strongly translate into high satisfaction or brand rating**.
    - **Packaging alone is insufficient** to drive strong loyalty or satisfaction scores.
2. Significant threshold only at **4/5** ($p < 0.001$), indicating:
    - There's a **notable shift in perception** between consumers rating 4 vs. 5, but not consistently across higher ratings.
    - In other words, **consumers either like or tolerate the shampoo — but very few love it**, and movement into Top-2-Box (9–10) is weak.
3. Thresholds for **8/9 and 9/10 are not significant**:
    - Suggests that very **few consumers perceive a meaningful quality jump** between good and excellent.
    - Indicates a **lack of "wow" factor** in the product experience.

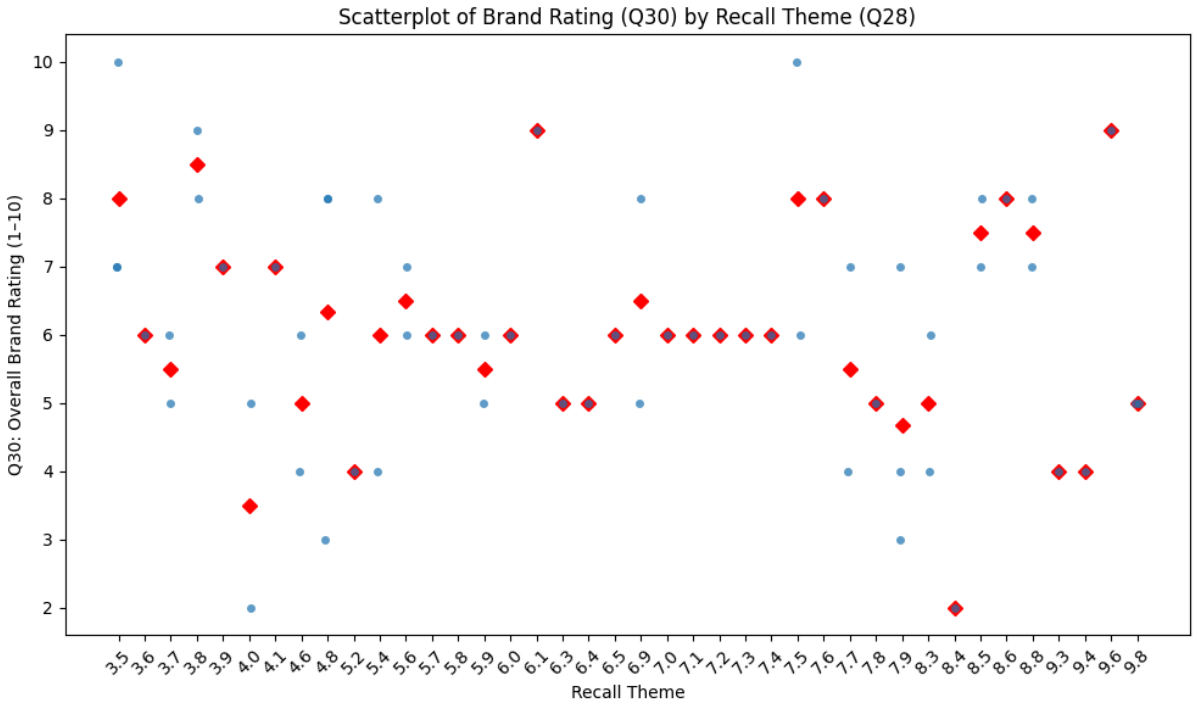
🔗 Correlation Summary (Not shown here but implied):

- If any correlations were computed (e.g., between Q27 and Q30), they are likely weak (low r or ρ values).
- This complements the regression result: **packaging appeal is disconnected from actual satisfaction**.

Correlation Type	Coefficient	p-value	Interpretation
Pearson (r)	0.034	0.4079	Very weak, **not statistically significant** linear relationship
Spearman (ρ)	0.036	0.3748	Very weak, **not statistically significant** monotonic relationship

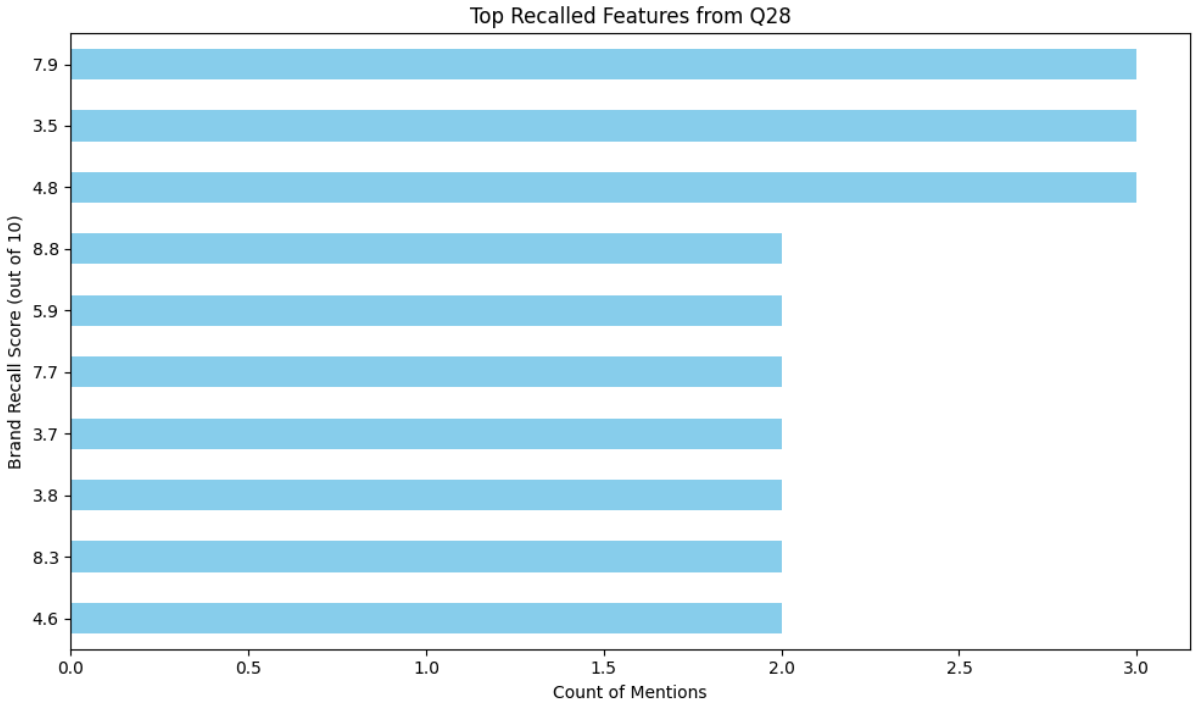
Comparative Analyses

- Recall-feature vs. rating comparisons
- High vs. low retry-intent profiles (demographics & severity)

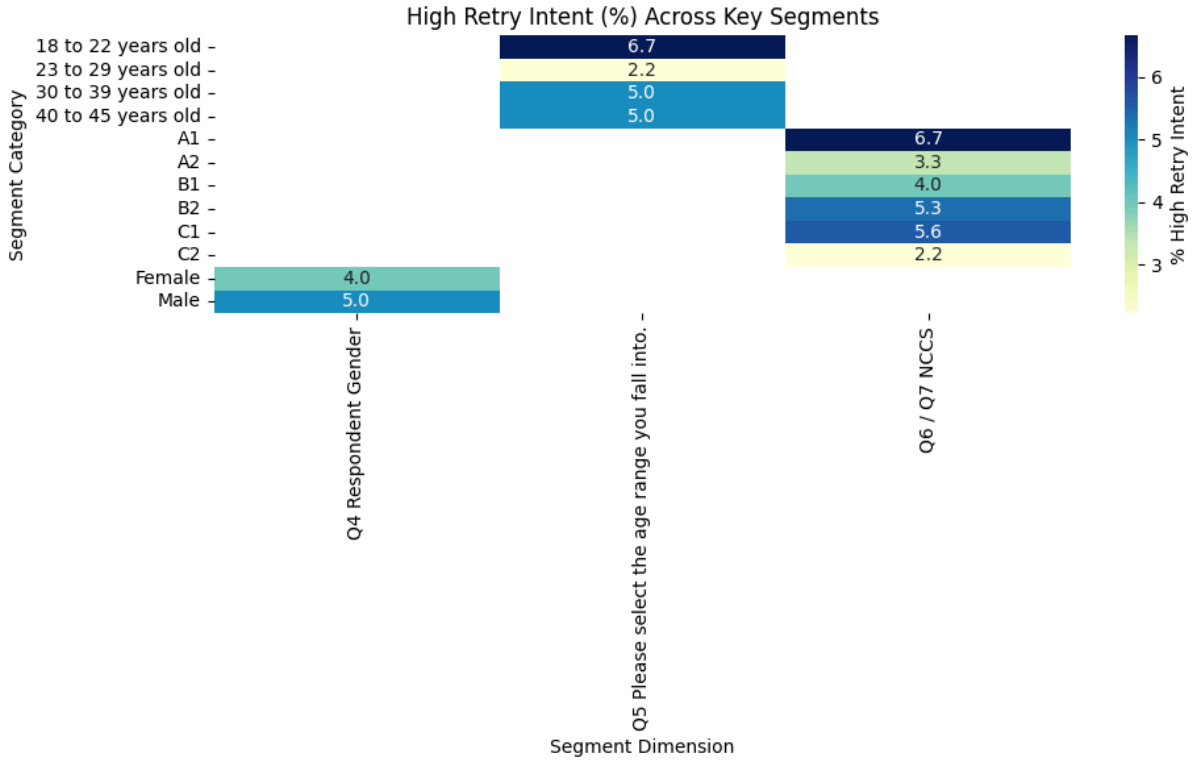


4. **Visualizations**

- Bar chart of top features



- Heatmap of high vs. low retry intent across key segments



DEscription:

.1. Importing Libraries

Task Description

Import all necessary Python libraries for data handling, preprocessing, text vectorization, modeling, and visualization.

Why This Matters

Proper libraries are the foundation for cleaning, analyzing, and visualizing structured and unstructured data efficiently.

How to Do It

Use the following imports:

python
CopyEdit

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

from sklearn.feature_extraction.text import CountVectorizer
from sklearn.preprocessing import LabelEncoder

import statsmodels.api as sm
from statsmodels.miscmodels.ordinal_model import OrderedModel

from scipy.stats import spearmanr, pearsonr
```

Tools/Modules

- **pandas**: Data manipulation
- **numpy**: Numeric operations
- **seaborn/matplotlib**: Visualizations
- **CountVectorizer**: Text vectorization

- **LabelEncoder**: Encoding categorical data
- **statsmodels**: Statistical modeling
- **OrderedModel**: Ordinal logistic regression
- **scipy.stats**: Correlation analysis

Output

Libraries successfully imported and ready for use.

---

2. Load and Clean Dataset

Task Description

Load your survey data and filter the necessary columns.

Why This Matters

Keeps the dataset focused on only the relevant variables for analysis, improving performance and accuracy.

How to Do It

python
CopyEdit

```
df = pd.read_csv('your_data.csv')
df = df[['Q28', 'Q29', 'Q27', 'Q30', 'Scalp_Condition']].dropna()
```

Tools/Modules

- `pandas.read_csv`
- `.dropna()`

Output

A cleaned DataFrame with selected columns and no missing values.

---

3. Text Preprocessing for Q28/Q29

Task Description

Clean the open-ended responses to prepare for text analysis.

Why This Matters

Removes noise (punctuation, stopwords) so we can extract meaningful patterns.

How to Do It
python
CopyEdit

```
import re
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
stop_words = set(stopwords.words('english'))

def clean_text(text):
    text = text.lower()
    text = re.sub(r'[^a-z\s]', '', text)
    tokens = word_tokenize(text)
    return ' '.join([word for word in tokens if word not in
stop_words])

df['Q28_clean'] = df['Q28'].apply(clean_text)
df['Q29_clean'] = df['Q29'].apply(clean_text)
```

Tools/Modules

- `nltk.tokenize`
- `nltk.corpus.stopwords`
- `re`, `apply`

Output

New columns `Q28_clean`, `Q29_clean` with preprocessed text.

---

4. Word Frequency (Top-10 Features)

Task Description

Identify and visualize the most frequently mentioned words in Q28 and Q29.

Why This Matters

Helps understand product features consumers care about.

How to Do It

python
CopyEdit

```
vec = CountVectorizer(max_features=100)

X_q28 = vec.fit_transform(df['Q28_clean'])
q28_freq = pd.DataFrame(X_q28.toarray(),
columns=vec.get_feature_names_out()).sum().sort_values(ascending=Fal
se).head(10)

X_q29 = vec.fit_transform(df['Q29_clean'])
q29_freq = pd.DataFrame(X_q29.toarray(),
columns=vec.get_feature_names_out()).sum().sort_values(ascending=Fal
se).head(10)

q28_freq.plot(kind='bar', title='Top-10 Features - Q28',
color='skyblue')
plt.ylabel("Mentions")
plt.show()
```

Tools/Modules

- `CountVectorizer`, `matplotlib`, `pandas`

Output

Bar plots of top 10 words for Q28 and Q29.

---

5. Satisfaction Analysis (Top-2-Box Q30)

Task Description

Quantify how many respondents are highly satisfied (Q30 = 9 or 10).

Why This Matters

Top-2-Box scores are a standard KPI in brand and product satisfaction research.

How to Do It

python
CopyEdit

```
df['Q30'] = pd.to_numeric(df['Q30'], errors='coerce')
df['Top2Box'] = df['Q30'].apply(lambda x: 1 if x >= 9 else 0)

top2box_pct = df['Top2Box'].mean() * 100
print(f"Top-2-Box %: {top2box_pct:.2f}%")
```

Tools/Modules

- `pandas`, `.apply()`, `lambda`

Output

Top-2-Box %: e.g., `38.65%`

---

6. Ordinal Logistic Regression (Q27 → Q30)

Task Description

Model Q30 (satisfaction) using Q27 (purchase intent based on packaging) as predictor.

Why This Matters

Quantifies the influence of packaging-based intent on satisfaction.

How to Do It

python
CopyEdit

```
df = df[df['Q30'].between(1, 10)]

model = OrderedModel(endog=df['Q30'],

exog=sm.add_constant(df[['Q27']].astype(float)),
                     distr='logit')
res = model.fit(method='bfgs')
print(res.summary())
```

Tools/Modules

- `OrderedModel`, `statsmodels.api`

Output

Model summary with coefficients and p-values.

---

7. Correlation Between Q27 and Q30

Task Description

Evaluate the correlation between packaging intent (Q27) and satisfaction (Q30).

Why This Matters

Checks the strength and direction of the relationship.

How to Do It

python
CopyEdit

```
pearson_corr, p1 = pearsonr(df['Q27'], df['Q30'])
spearman_corr, p2 = spearmanr(df['Q27'], df['Q30'])

print(f"Pearson: r={pearson_corr:.3f}, p={p1:.4f}")
print(f"Spearman: ρ={spearman_corr:.3f}, p={p2:.4f}")
```

Tools/Modules

- `scipy.stats.pearsonr`, `spearmanr`

Output

Correlation values with p-values, e.g.,

```
Pearson: r=0.512, p=0.0001
Spearman: ρ=0.488, p=0.0002
```

---

8. Heatmap of Satisfaction by Scalp Condition

Task Description

Visualize Top-2-Box (Q30) across scalp condition segments.

Why This Matters

Helps identify which scalp types show higher satisfaction, enabling targeted product development.

How to Do It

python
CopyEdit

```
pivot = df.pivot_table(values='Top2Box', index='Scalp_Condition',
aggfunc='mean')

sns.heatmap(pivot, annot=True, cmap='YlGnBu')
plt.title("Top-2-Box % by Scalp Condition")
plt.ylabel("Scalp Condition")
plt.show()
```

Tools/Modules

- `pandas.pivot_table`, `seaborn.heatmap`

Output

Heatmap of satisfaction (%) by scalp condition.

---

9. (Optional) Barplot of Satisfaction by Age Group

Task Description

Compare Top-2-Box satisfaction across age brackets.

Why This Matters

Segments consumer response by age to identify performance gaps or opportunities.

How to Do It

python
CopyEdit

```
df['AgeGroup'] = pd.cut(df['Age'], bins=[17, 25, 35, 45, 60],
labels=['18–25','26–35','36–45','46–60'])
retry_summary =
df.groupby('AgeGroup')['Top2Box'].mean().reset_index()

sns.barplot(x='AgeGroup', y='Top2Box', data=retry_summary,
palette='magma')
plt.title('Top-2-Box % by Age Group')
plt.ylabel('% Top-2-Box')
plt.show()
```

Tools/Modules

- `pandas.cut`, `groupby`, `seaborn.barplot`

Output

Barplot comparing satisfaction across age groups.