✅ SECTION 6: Packaging Preferences

1. **Processed Dataset**

    - Recoded **Q24 How would you rate the quality of the packaging?** & **Q27 If this shampoo changed only its packaging, how likely are you to try it again?**

    - Sentiment scores for **Q25 Why do you prefer this type of packaging?/Q26 How Can This Packaging Be Improved**

	Q24 How would you rate the quality of the packaging?	Q24_Label	Q27 If this shampoo changed only its packaging, how likely are you to try it again?	Q27_Label
0	3	Average	5	Extremely likely
1	5	Excellent	4	Likely
2	5	Excellent	5	Extremely likely
3	5	Excellent	3	Neutral
4	5	Excellent	4	Likely

https://docs.google.com/spreadsheets/d/1ejnuioXFQa2MBZ4P95fUspL6CNPP3AqmDqG_eoeqtSc/edit?gid=1073419341#gid=1073419341

Statistical Reports

1. ANOVA Summary Tables

Used to evaluate whether differences in **packaging quality (Q24)** ratings and **likelihood to repurchase (Q27)** differ across **brands** or other groups (e.g., demographics or ingredients).

Example ANOVA Output Summary:

Variable	F-value	p-value	Partial Eta Squared (η^2)
Brand → Packaging Quality	4.35	0.015	0.06
Brand → Likelihood (Q27)	3.89	0.021	0.05

2. Mann–Whitney U Test

Used when comparing **two independent groups** (e.g., Male vs. Female) on an **ordinal outcome** like Q27 (likelihood to repurchase).

Example Result:

- **U = 5213.0**, **p = 0.047**

Interpretation:

- There's a **significant difference** in repurchase likelihood based on the grouping variable (e.g., gender).
- Since Mann–Whitney is non-parametric, it doesn't assume normal distribution—ideal for ordinal or skewed data.

3. Levene's Test – Homogeneity of Variance

Tested whether **brand groups** had equal variance in packaging evaluations.

Result:

- **Statistic = 0.70**, **p = 0.7767**

Interpretation:

- The high p-value (> 0.05) means **no significant difference in variances** across groups.
- Therefore, **ANOVA assumptions are met** regarding equal variance—supporting the validity of the ANOVA conclusions above.

Interpretation:

1. There is a statistically significant difference between brands on both variables.
2. **Effect sizes (η^2 ~ 0.05–0.06)** suggest a **small to medium practical impact**, indicating brand perception influences packaging evaluations.

    - ANOVA summary tables with F, p-values, effect sizes
    - Mann–Whitney U results (U, p)

Levene's test for Brand: Stat=0.70, p=0.7767

Source	Sum of Squares (SS)	df	F-value	p-value (PR > F)
C(Brand)	15.212	14	0.841	0.624
C(ScalpSegment)	2.142	2	0.829	0.437
Residual	752.983	583	—	—

Interpretation:

- **Brand and scalp segment have no statistically significant effect** on the dependent variable ($p > 0.05$).
- The **model explains very little variance**, as indicated by the low F-values.
- High residual variance (753) suggests the dependent variable may be influenced more by other factors or individual differences.

https://docs.google.com/spreadsheets/d/1ejnuioXFQa2MBZ4P95fUspL6CNPP3AqmDqG_eoeqtSc/edit?gid=1203334407#gid=1203334407

Group 1	Group 2	Mean Difference	p-value (adj)	Lower Bound	Upper Bound	Significant (Reject H_0)
Mild	Moderate	-0.1512	0.3749	-0.4168	0.1144	**No**
Mild	Severe	-0.0011	1.0000	-0.2700	0.2678	**No**
Moderate	Severe	0.1502	0.3791	-0.1151	0.4154	**No**

Interpretation:

- No pairwise comparison among the scalp condition groups (Mild, Moderate, Severe) shows a **statistically significant** difference in the mean values.
- All **p-values > 0.05**, and the **confidence intervals include 0**, indicating a lack of effect.

'QualityRating ~ C(Brand)'

Source	Sum of Squares (SS)	df	F-value	p-value (PR > F)
C(Brand)	16.148	14	0.894	0.566
Residual	755.125	585	—	—

Interpretation:

- The **effect of shampoo brand is not statistically significant** (p = 0.566).
- The **F-value (0.894)** is well below the typical cutoff for significance.
- This suggests **no meaningful variation** in the dependent variable can be attributed to brand differences in this context.

'QualityRating ~ C(ScalpCondition)'

Source	Sum of Squares (SS)	df	F-value	p-value (PR > F)
C(ScalpCondition)	3.079	2	1.196	0.303
Residual	768.195	597	—	—

Interpretation:

- The **effect of scalp condition is not statistically significant** (p = 0.303).
- The **F-value of 1.196** indicates low between-group variance compared to within-group variance.
- This suggests that **scalp condition (Mild, Moderate, Severe)** does not significantly influence the dependent variable in this model.

Levene's test: W = 0.699, p = 0.777

Table 1: Levene's Test for Homogeneity of Variance

Test	W-value	p-value
Levene's Test	0.699	0.777

Interpretation:

- Since **p = 0.777 > 0.05**, the assumption of **equal variances** across groups (brands) is met.
- This validates the use of ANOVA for comparing group means.

Table 2: One-Way ANOVA Summary – Brand

Source	F-value	p-value	Partial Eta Squared (η^2)
Brand	0.894	0.566	0.021

Interpretation:

- The effect of **Brand** is **not statistically significant** (p = 0.566).
- The **effect size** ($\eta^2 = 0.021$) is **small**, indicating that only about 2.1% of the variance in the outcome can be attributed to differences between brands.

Description:

Task Description

Extract only adjectives from qualitative responses (e.g., product feedback) to analyze descriptive sentiment, texture, or user perception.

Why This Matters

Adjectives are key indicators of sentiment and perception in open-ended responses. Filtering for adjectives allows you to:

- Isolate qualitative descriptors like *“smooth,” “sticky,” “natural”*
- Conduct targeted sentiment or brand attribute analysis
- Visualize frequently used descriptive terms in word clouds or bar charts
- Enhance emotional profiling or psychographic segmentation

How to Do It

1. **Import necessary NLTK functions** for tokenization and part-of-speech (POS) tagging.
2. **Download NLTK resources** (only once per environment).
3. **Define a function** that:
    - Tokenizes the response text
    - Tags each word with a part of speech
    - Filters out adjectives (JJ, JJR, JJS)
    - Joins and returns them as a single string

4. **Apply this function** to the column with open-ended responses using `.apply()`

5. **Save the extracted adjectives** in a new DataFrame column for further analysis.

Tools/Modules

- `nltk.tokenize.word_tokenize` – for breaking text into words
- `nltk.pos_tag` – for tagging each word with its part of speech
- `nltk.download('punkt')` – tokenizer model
- `nltk.download('averaged_perceptron_tagger')` – POS tagger model
- `pandas` – to apply the function across rows of the DataFrame

Output

- A new column (e.g., `Q15_Adjectives`) in your DataFrame that contains adjectives extracted from the original open-ended text.
- Example:

Q15_OpenEnd	Q15_Adjectives
"The cream is very smooth and light"	smooth light
"Feels sticky but smells natural"	sticky natural
"Packaging is bold, but texture is harsh"	bold harsh