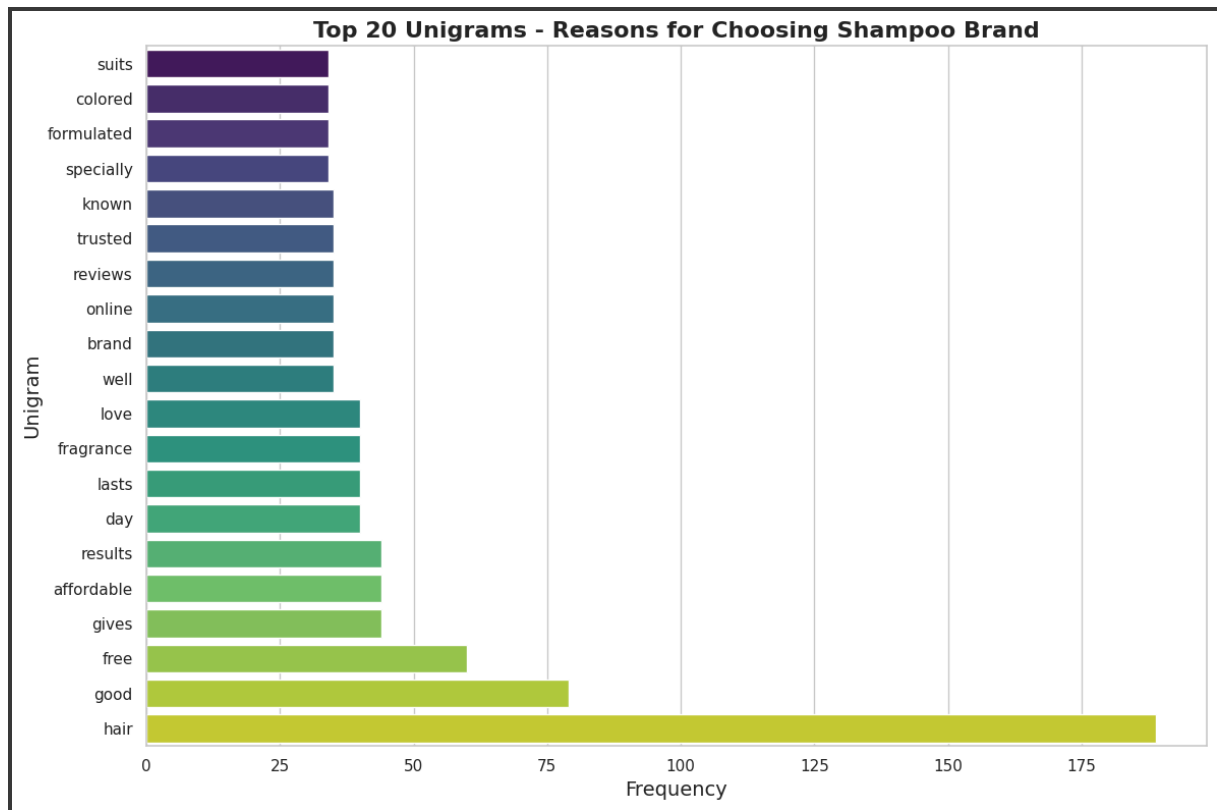


## SECTION 5: Brand & Ingredient Choices

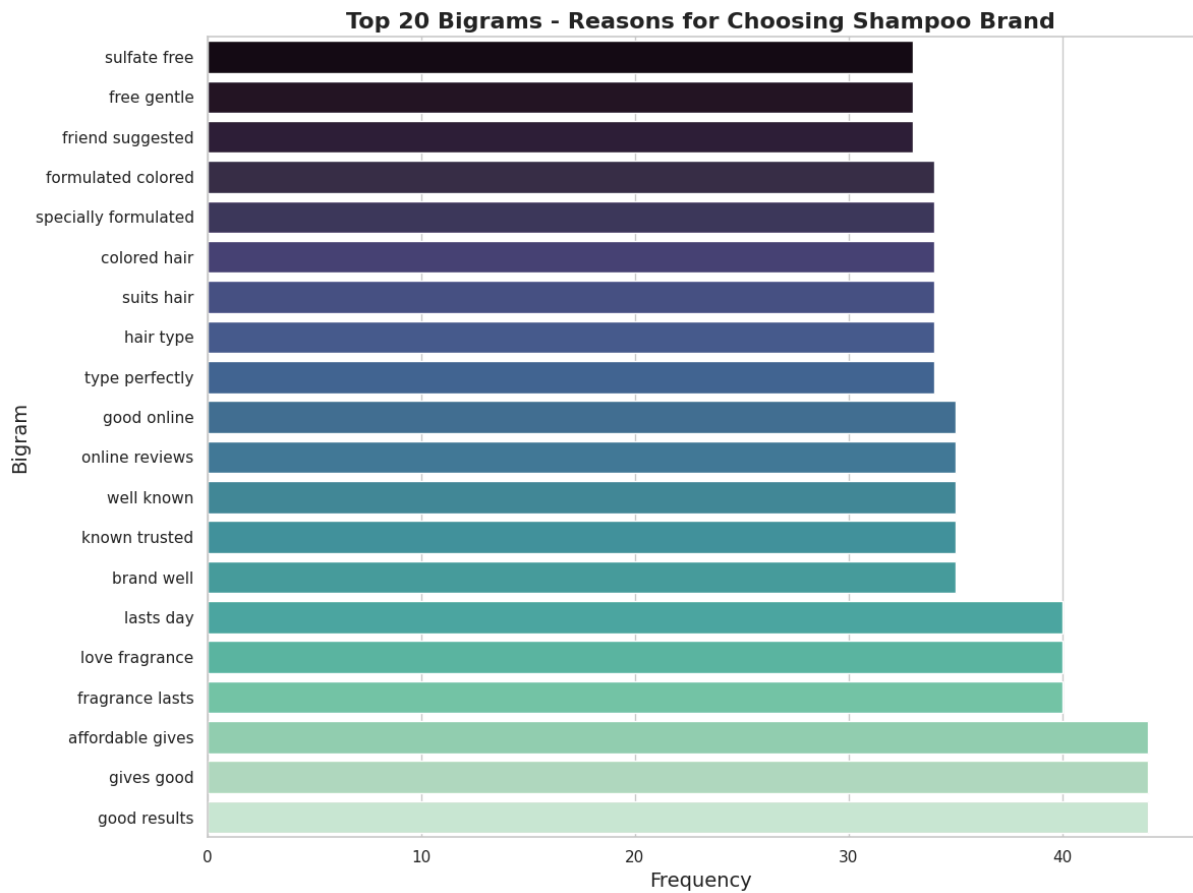
### Unigram and Bigram of

Q19: “Why did you choose your current shampoo brand?”



### Unigrams – Top Keywords

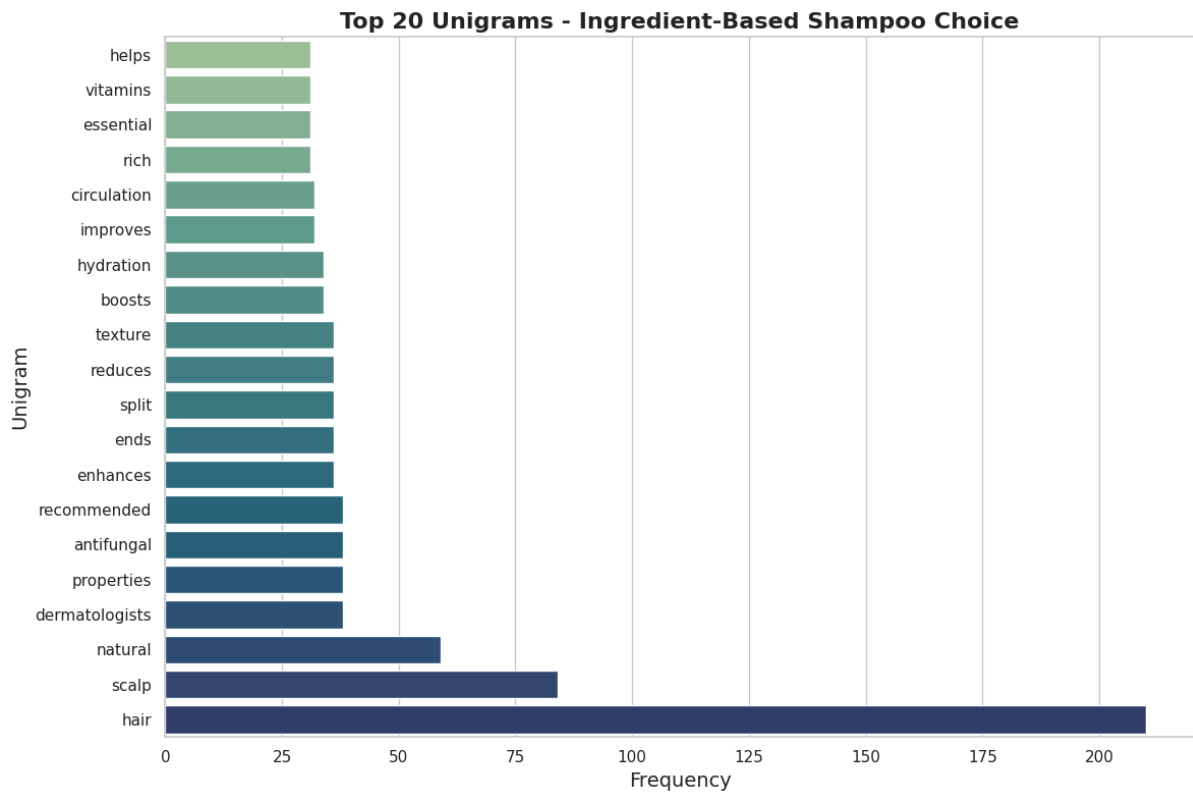
- **“Hair”** was the most mentioned term, confirming product relevance and direct functionality.
- Words like **“affordable”**, **“good”**, and **“results”** highlight practical value perceptions.
- Emotional and sensory drivers are reflected in **“love”**, **“fragrance”**, and **“trusted”**.



### Bigrams – Key Phrases

- Frequent value chains: **“affordable gives good results”**, **“love fragrance lasts day”**.
- Sentiment and trust revealed in **“friend suggested”** and **“online reviews”**.
- Needs-based language appears in **“specially formulated”**, **“colored hair”**, **“sulfate free”**.

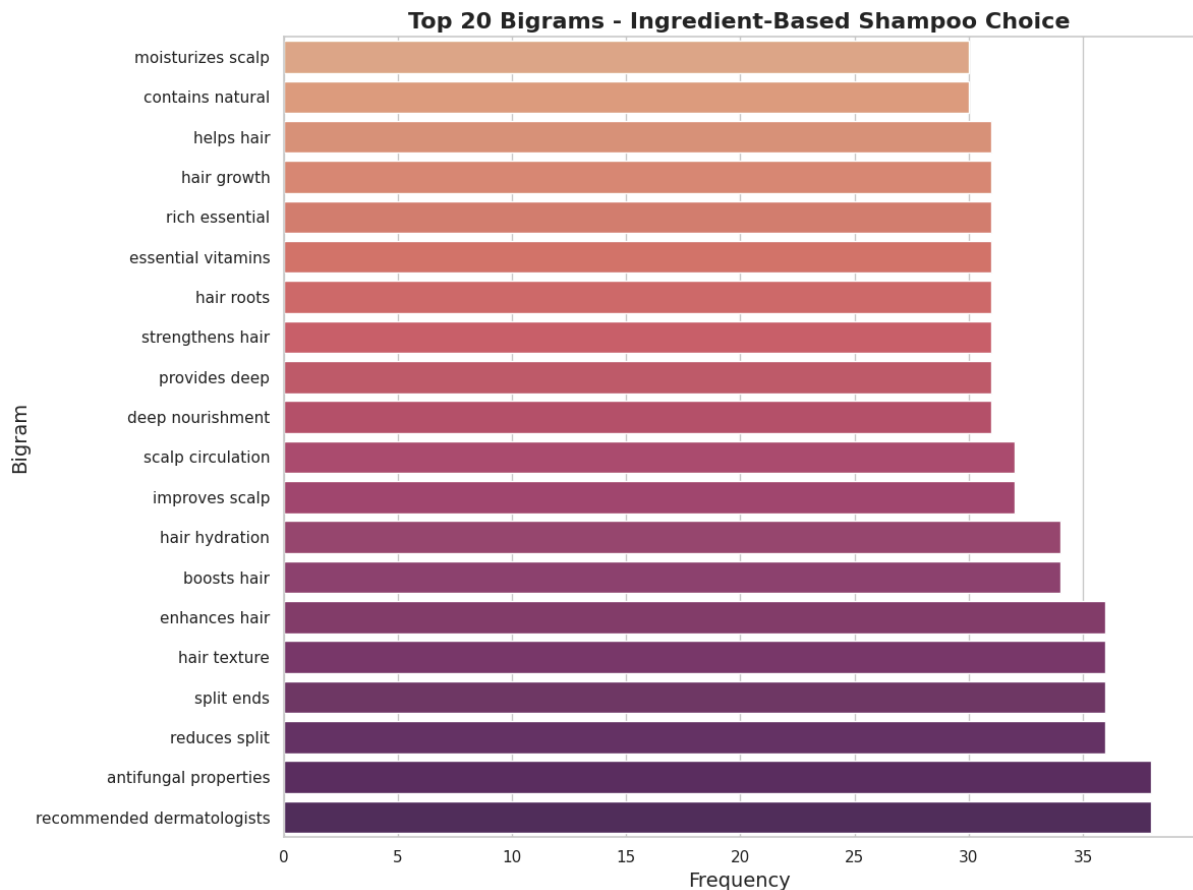
Q20: “Why did you choose shampoos with these ingredients?”



### Top Unigrams (Individual Words)

This bar chart reveals the most commonly used individual words when consumers describe why they chose shampoos with specific ingredients.

- **“Hair”** and **“Scalp”** dominate, suggesting primary concern areas.
- Functionally beneficial words like **“antifungal”**, **“hydration”**, and **“reduces”** show that consumers are health- and benefit-conscious.
- Presence of **“recommended”** and **“dermatologists”** highlights the influence of expert endorsements.
- Keywords such as **“texture”**, **“nourishment”**, and **“essential vitamins”** suggest a desire for improved appearance and long-term care.

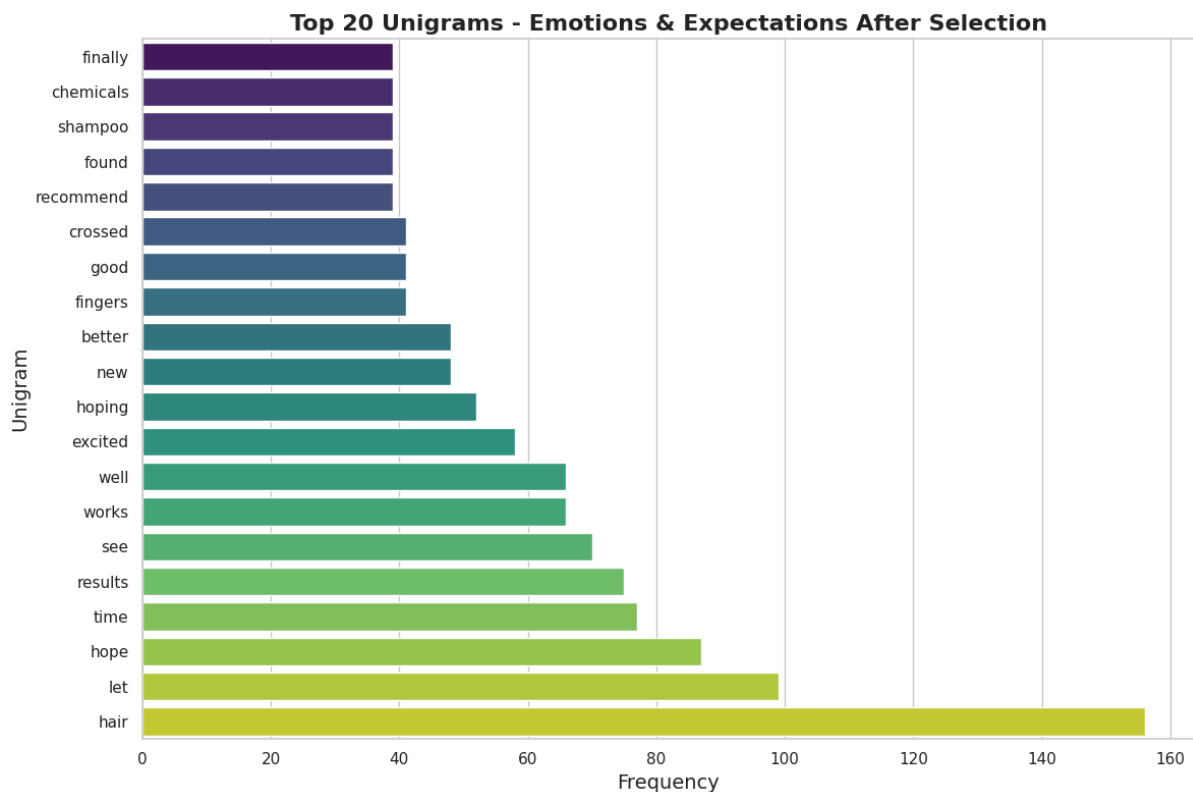


### Top Bigrams (Two-Word Phrases)

This chart provides insight into the exact phrases used, giving context and depth to ingredient preferences.

- Key benefits appear in phrases such as **“antifungal properties”**, **“deep nourishment”**, and **“moisturizes scalp”**, confirming functional value perceptions.
- Multiple health and efficacy indicators: **“hair growth”**, **“reduces split ends”**, **“strengthens hair”**, **“improves scalp circulation”**.
- Expert and trust-driven drivers are evident in **“recommended dermatologists”**.
- Natural appeal is captured through **“contains natural”** and **“rich essential vitamins”**.

Q21: “What thought or feeling did you have after selecting your current shampoo?”

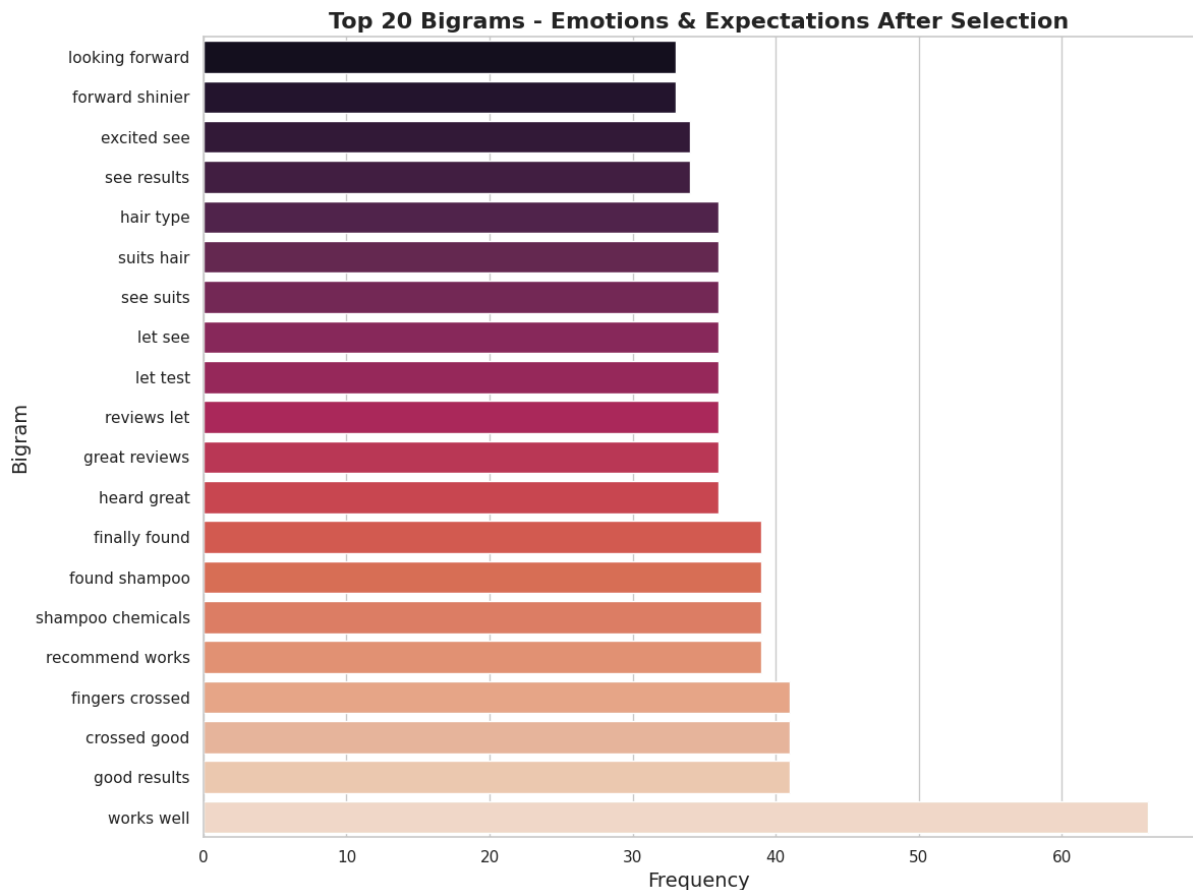


### Top Unigrams (Single Words)

This chart highlights the most frequent **emotional and cognitive responses** from consumers after choosing their shampoo:

- **“Hair,” “results,” and “shampoo”** remain functional keywords—still the focus.
- Emotional states and optimism dominate with **“hope,” “hoping,” “excited,” “finally,” and “fingers crossed.”**
- **“Let,” “see,” “works,” “better”** reflect a trial mindset — suggesting users are *testing* performance.
- **“Recommend,” “chemicals,” and “new”** indicate both trust-building and caution, especially related to ingredient concerns.


**Key Insight:** Consumers often approach their purchase with **hopeful anticipation**, showing signs of **uncertainty, cautious optimism**, and **expectation of transformation**.

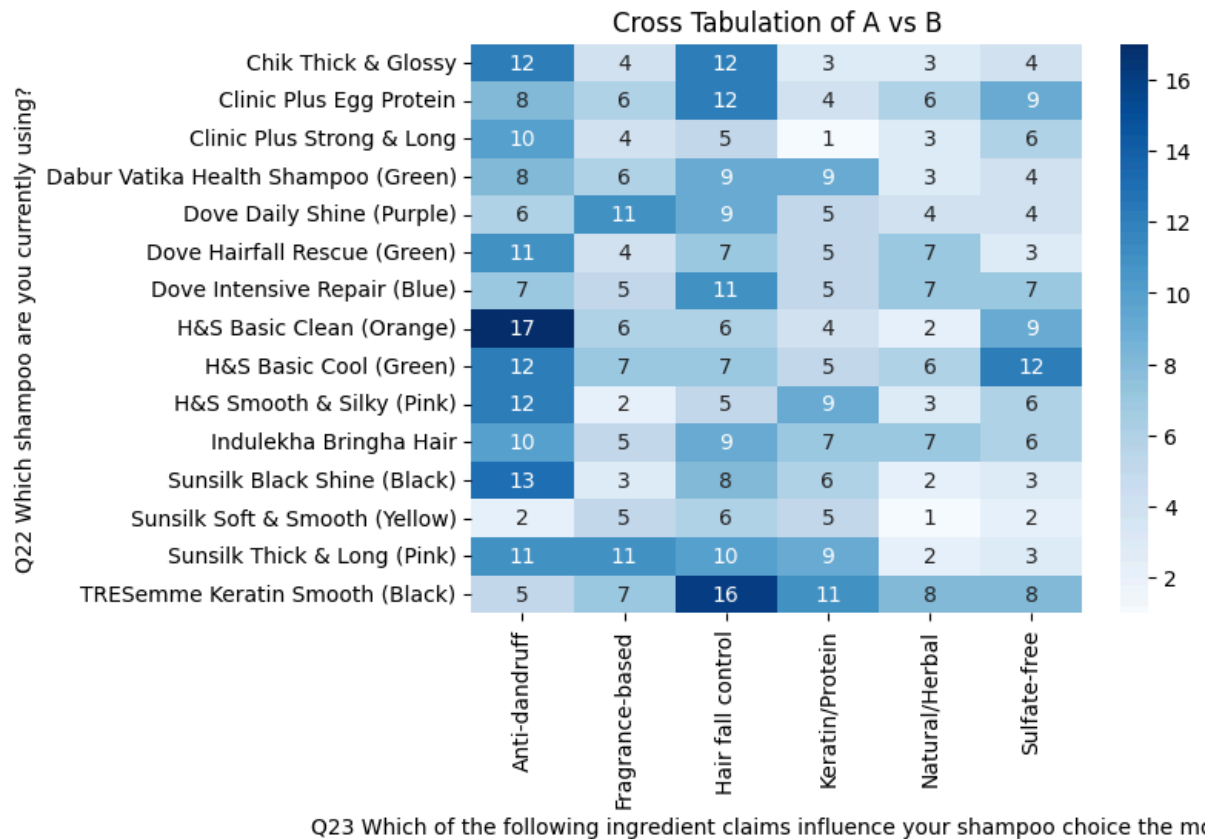


### Top Bigrams (Two-Word Phrases)

This plot offers deeper **contextual phrases** that reflect emotional undercurrents:

- **“Fingers crossed,” “good results,” and “finally found”** show relief and cautious optimism.
- Phrases like **“works well,” “see results,” “excited see”** reinforce a **wait-and-watch** attitude.
- **“Shampoo chemicals,” “recommend works”** reflect **ingredient sensitivity** and validation-driven decisions.
- Sentences such as **“heard great,” “great reviews,” “let test”** suggest **word-of-mouth and research impact**.
- **“Looking forward” and “forward shinier”** reflect aspirational goals related to beauty and confidence.

 **Key Insight:** The purchase is followed by a **hopeful, evaluative mindset**, driven by expectations of **visible improvement** and **external validation** (reviews, recommendations).



Optimization terminated successfully.

Current function value: 0.517588

Iterations 6

	Odds Ratio	p-value	CI Lower	CI Upper
const	0.462	0.027	0.233	0.915
Clinic Plus Egg Protein	0.788	0.623	0.304	2.039
Clinic Plus Strong & Long	0.451	0.187	0.138	1.471
Dabur Vatika Health Shampoo (Green)	0.650	0.404	0.236	1.787
Dove Daily Shine (Purple)	0.650	0.404	0.236	1.787
Dove Hairfall Rescue (Green)	0.506	0.211	0.173	1.474
Dove Intensive Repair (Blue)	0.769	0.595	0.291	2.028
H&S Basic Clean (Orange)	0.342	0.056	0.114	1.027
H&S Basic Cool (Green)	0.361	0.058	0.126	1.035

H&S Smooth & Silky (Pink)	0.339	0.068	0.106	1.085
Indulekha Bringha Hair	0.557	0.253	0.205	1.518
Sunsilk Black Shine (Black)	0.642	0.405	0.226	1.824
Sunsilk Soft & Smooth (Yellow)	0.867	0.810	0.270	2.787
Sunsilk Thick & Long (Pink)	0.602	0.309	0.226	1.602
TRESemme Keratin Smooth (Black)	0.889	0.797	0.362	2.182

[https://docs.google.com/spreadsheets/d/1ejnuioXFQa2MBZ4P95fUspL6CNPP3AqmDqG\\_eoeqtSc/edit?gid=1003275239#gid=1003275239](https://docs.google.com/spreadsheets/d/1ejnuioXFQa2MBZ4P95fUspL6CNPP3AqmDqG_eoeqtSc/edit?gid=1003275239#gid=1003275239)

---

## Description:

### Task Description

Extract adjectives from open-ended text responses using natural language processing to understand descriptive language used by consumers.

---

### Why This Matters

Adjectives reveal consumer sentiment, product descriptors, and emotional associations. This helps in brand language analysis, product feedback interpretation, and emotional positioning strategies.

---

### How to Do It

1. **Import required libraries** (`nltk`, `pandas`) and download necessary models.
2. **Define a function** that:
  - Handles missing values.
  - Tokenizes the input text into words.
  - Assigns part-of-speech (POS) tags to each token.



- Filters only adjectives (**JJ**, **JJR**, **JJS**).
  - Joins them into a clean string.
3. **Apply the function** to the relevant open-ended column in your DataFrame.
- 

## Tools/Modules

- **nltk** (Natural Language Toolkit)
  - **word\_tokenize** (for splitting text)
  - **pos\_tag** (for POS tagging)
  - **pandas** (for DataFrame handling)
- 

## Output

A new DataFrame column (e.g., **Q15\_Adjectives**) containing only the adjectives extracted from each open-ended response.

---

## Code

```
python
CopyEdit
import pandas as pd
import nltk
from nltk.tokenize import word_tokenize
from nltk import pos_tag

# Download NLTK models (run once)
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')

# Define adjective extraction function
def extract_adjectives(text):
    if pd.isnull(text):
        return ""
```

```
    tokens = word_tokenize(text)
    tagged = pos_tag(tokens)
    adjectives = [word for word, tag in tagged if tag in ['JJ',
'JJR', 'JJS']]
    return ' '.join(adjectives)

# Apply to open-ended responses
df['Q15_Adjectives'] = df['Q15_OpenEnd'].apply(extract_adjectives)
```