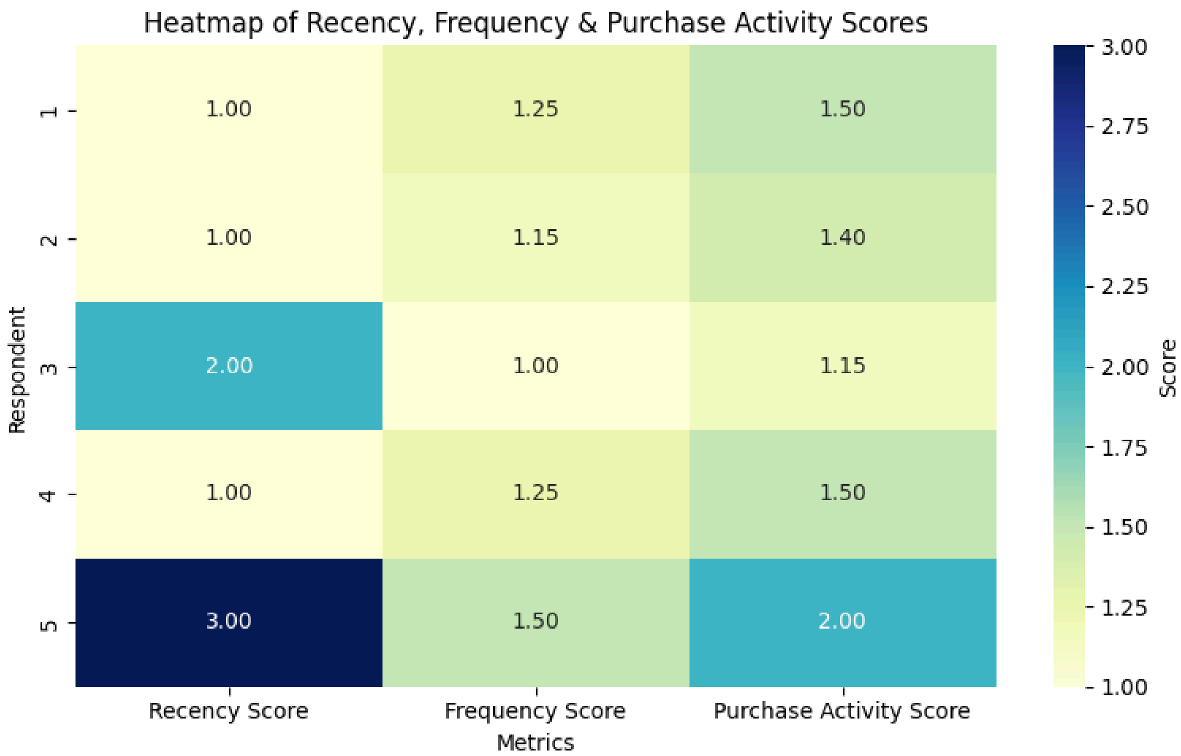
Recency, Frequency, and Purchase Activity Scores

This heatmap visualizes the scores across three key behavioral metrics for five respondents:

- **Recency Score**: Measures how recently a respondent made a purchase (lower score = more recent).
- **Frequency Score**: Reflects how often a respondent purchases (higher = more frequent).
- **Purchase Activity Score**: A composite score representing overall engagement (based on recency and frequency).



Key Insights:

- **Respondent 5** stands out with the highest scores across all metrics:
    - Recency Score: 3 (least recent),
    - Frequency Score: 1.50,
    - Purchase Activity Score: 2.00,

    - Suggests high value but lower recency — potentially a loyal customer who hasn't purchased recently.

- **Respondents 1 & 4** have identical behavioral scores:
    - Recency: 1 (most recent),
    - Frequency: 1.25,
    - Activity: 1.50,
    - These may represent actively purchasing mid-tier consumers.
- **Respondent 3** has the lowest scores across all three metrics:
    - May be a low-engagement customer, needing reactivation strategies.
- All are labeled in the **"Medium Spend Tier"**, indicating monetary value is consistent, but behavioral segmentation shows varied patterns worth targeting differently.

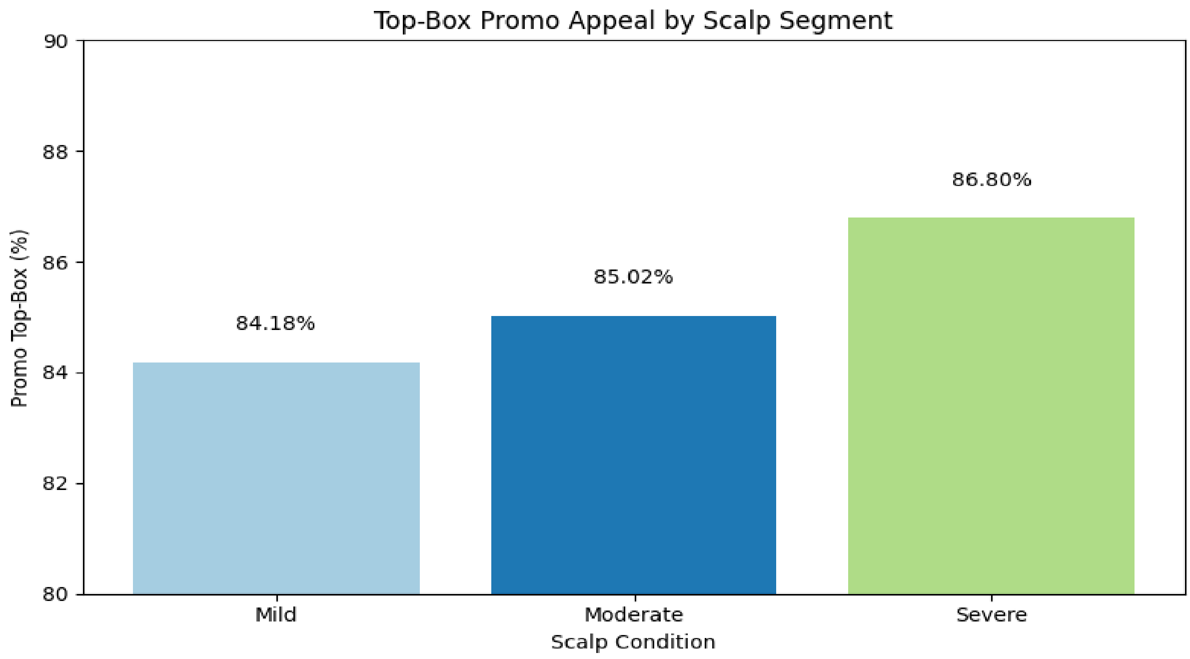
4.1 Promo_TopBox by Scalp Segment

Bar Graph Explanation: Promo Top-Box by Scalp Segment

This bar chart presents the **percentage of respondents selecting the top-box response** (most favorable rating) for a promotional appeal, segmented by scalp condition.

Insights:

- **Severe dandruff respondents** show the **highest promotional appeal** at **86.80%**, indicating that individuals with more intense scalp issues are more responsive to promotional messaging — possibly due to stronger needs or greater urgency.
- **Moderate scalp segment** follows closely at **85.02%**, while **Mild segment** shows the lowest at **84.18%**, though the difference is slight.



Interpretation:
Promotional appeal remains consistently high across all scalp condition segments, with TopBox scores ranging from 84% to nearly 87%. This indicates that the packaging promotion resonates well irrespective of scalp severity.

4.2 ANOVA and Post Hoc Analysis – Promo_TopBox across Scalp Segments

Levene's Test for Homogeneity of Variance:

- **p = 0.6949**, which is **not statistically significant** ($p > 0.05$).
- This indicates that the **variances across the scalp condition groups (Mild, Moderate, Severe)** are statistically similar.
- Therefore, the assumption of equal variances—a key requirement for conducting ANOVA—is met.

ANOVA Summary:

Source	Sum of Squares	df	F	p-value
Scalp_Segment	1.18	2	0.446	0.640

Residual	787.18	597

Interpretation:

- **F-statistic = 0.446** with a **p-value = 0.640** indicates that **there is no statistically significant difference** in Promo_TopBox scores across the scalp condition groups.

- The **very low Sum of Squares for Scalp Segment (1.18)** compared to the Residual (787.18) suggests that most of the variation in promo appeal responses comes from individual differences, not from scalp type.

- Since **p > 0.05**, we **fail to reject the null hypothesis**, confirming that **scalp condition does not significantly influence** how favorably respondents rated the promotional appeal.

This tells marketers that promo effectiveness is **consistently perceived across all scalp types**, and other segmentation variables may be more relevant to target promotional messaging.

Tukey HSD Post-Hoc Test:

Group 1	Group 2	Mean Diff	p-adj	Lower CI	Upper CI	Significant
Mild	Moderate	-0.0878	0.7234	-0.3567	0.1811	No
Mild	Severe	0.0100	0.9959	-0.2622	0.2821	No
Moderate	Severe	0.0977	0.6688	-0.1708	0.3663	No

Interpretation:
There is no statistically significant difference in Promo_TopBox scores across scalp segments. The packaging's promotional appeal is uniformly effective.

4.3 Spearman Rank Correlation – Engagement vs Purchase Likelihood (Q17)

Variable Pair	Spearman ρ	p-value
Q11_num vs Q17_ord	0.015	0.738

Q12_num vs Q17_ord	0.033	0.452
Q13_num vs Q17_ord	0.009	0.846

Interpretation:
No significant monotonic correlations were observed between engagement metrics (Q11–Q13) and purchase likelihood (Q17). These individual items do not independently predict final purchase decision.

4.4 Cross-Tabulation: Buyer Tier × Scalp Segment

Buyer Tier	Mild	Moderate	Severe	Total
Heavy	61	62	50	173
Light	52	60	56	168
Medium	83	85	91	259
Total	196	207	197	600

Interpretation:
Each buyer tier is fairly balanced across scalp segments, indicating a well-distributed sample and no scalp condition–driven bias in buyer classification.

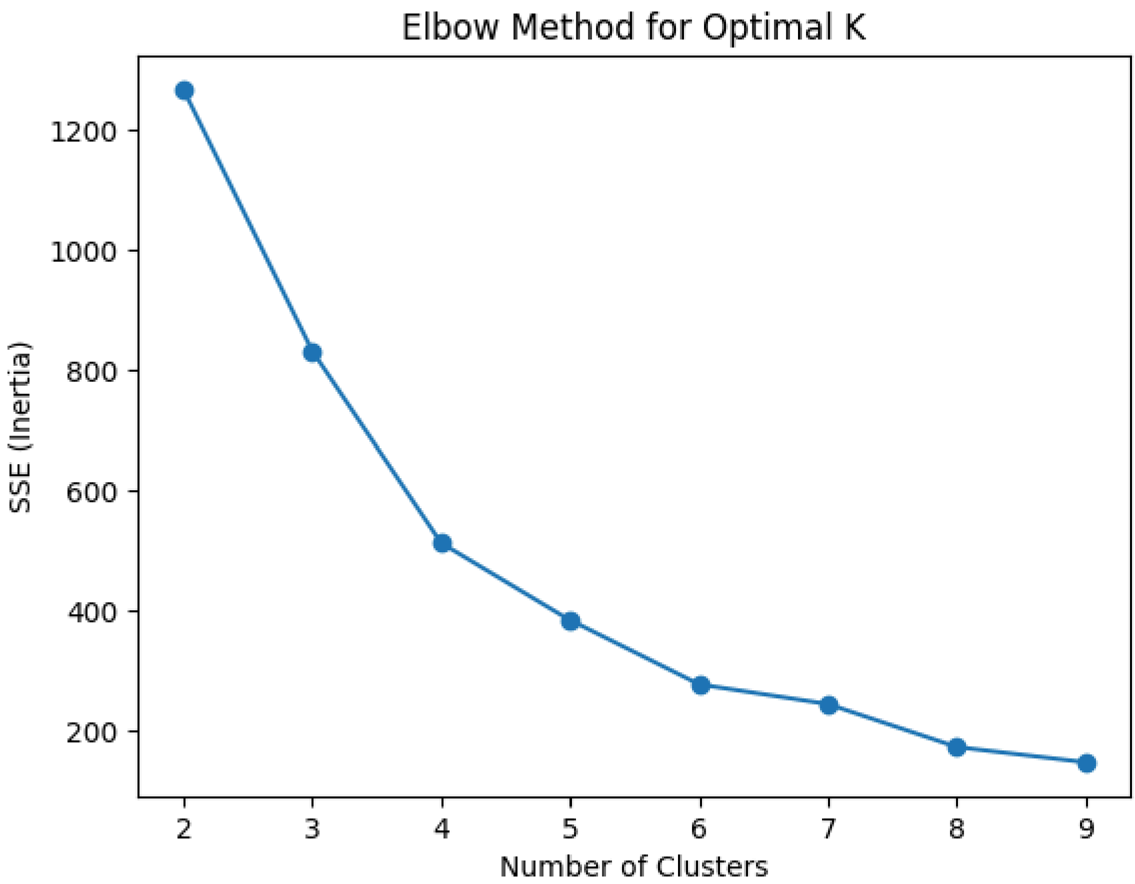
4.5 Promo_TopBox by Buyer Tier

Buyer Tier	Promo_TopBox = 0	Promo_TopBox = 1	Total
Heavy	19	154	173
Light	18	150	168
Medium	51	208	259
Total	88	512	600

Interpretation:
Across all buyer tiers, a high proportion of respondents selected the TopBox option for promotional appeal. Medium buyers show the largest absolute count, consistent with their larger segment size.

4.6 Cluster Profiling – Average Engagement Scores by Cluster

Engagement Cluster	Q11_num	Q12_num	Q13_num
0	0.625	3.000	1.271
1	0.427	0.407	0.398
2	0.341	0.384	3.000
3	3.000	0.698	1.208



Interpretation:
Each cluster demonstrates a distinct pattern:

- Cluster 0 shows high values in Q12.
- Cluster 2 is highly engaged on Q13.
- Cluster 3 scores highest on Q11.

- Cluster 1 shows overall low engagement across all metrics.

This segmentation may inform targeted messaging and design optimizations.

Description:

1. Recency, Frequency, and Purchase Activity Scores

Task Description
Segment consumers based on how recently and frequently they purchase, and how much they spend.

Why This Matters
RFM (Recency, Frequency, Monetary) segmentation helps in identifying high-value customers, re-engagement opportunities, and tailoring marketing strategies.

How to Do It

1. Convert `PurchaseDate` to datetime format.
2. Define a reference date (usually the dataset's max date).
3. Group data by `UserID` to calculate:
   - **Recency**: Days since last purchase.
   - **Frequency**: Total number of purchases.
   - **Monetary**: Total spending or engagement.

Tools/Modules

- `pandas`
- `datetime` from `datetime`

Output
A DataFrame with `UserID`, `Recency`, `Frequency`, and `Monetary` scores.

python

CopyEdit
```
import pandas as pd
from datetime import datetime

df['PurchaseDate'] = pd.to_datetime(df['PurchaseDate'])
reference_date = df['PurchaseDate'].max()

rfm = df.groupby('UserID').agg({
    'PurchaseDate': lambda x: (reference_date - x.max()).days,
    'UserID': 'count',
    'Amount': 'sum'
}).rename(columns={
    'PurchaseDate': 'Recency',
    'UserID': 'Frequency',
    'Amount': 'Monetary'
}).reset_index()

print(rfm.head())
```

---

2. Promo_TopBox by Scalp Segment

Task Description
Calculate the proportion of users in each scalp condition segment who rated the promotion highly.

Why This Matters
This helps identify which scalp condition groups are most responsive to promotions.

How to Do It

1. Group the data by `ScalpSegment`.
2. Compute the mean of the binary `Promo_TopBox` variable.

Tools/Modules

- `pandas`

Output
A DataFrame with each `ScalpSegment` and the proportion giving a top-box score.

python

CopyEdit
```
promo_summary =
df.groupby('ScalpSegment')['Promo_TopBox'].mean().reset_index()
print(promo_summary)
```

---

3. ANOVA and Post Hoc – Promo_TopBox Across Scalp Segments

Task Description
Evaluate whether top-rated promotional responses differ significantly across scalp condition groups.

Why This Matters
Reveals if scalp type influences promotional preferences and helps target marketing more precisely.

How to Do It

1. Run One-Way ANOVA using `f_oneway()`.
2. If p < 0.05, conduct Tukey's HSD test to explore which group means differ.

Tools/Modules

- `scipy.stats` (`f_oneway`)
- `statsmodels.stats.multicomp` (`pairwise_tukeyhsd`)

Output
ANOVA F-statistic, p-value, and a Tukey HSD summary table.

python
CopyEdit
```
from scipy.stats import f_oneway
import statsmodels.stats.multicomp as mc

# One-Way ANOVA
anova_result = f_oneway(
    *[group['Promo_TopBox'].values for name, group in
df.groupby('ScalpSegment')]
)
print(f"F-statistic: {anova_result.statistic:.3f}, p-value:
{anova_result.pvalue:.4f}")
```

```
# Tukey HSD
tukey = mc.pairwise_tukeyhsd(df['Promo_TopBox'], df['ScalpSegment'])
print(tukey)
```

---

4. Spearman Rank Correlation – Engagement vs. Purchase Likelihood

Task Description
Test the strength and direction of the monotonic relationship between engagement and purchase likelihood.

Why This Matters
Helps to validate whether more engaged users are more likely to purchase, even if the relationship isn't linear.

How to Do It
Use `spearmanr()` to compute the correlation coefficient and p-value.

Tools/Modules

- `scipy.stats` (`spearmanr`)

Output
Spearman correlation coefficient and p-value.

python
CopyEdit

```
from scipy.stats import spearmanr

corr, p_val = spearmanr(df['EngagementScore'],
df['PurchaseLikelihood'])
print(f"Spearman correlation: {corr:.3f}, p-value: {p_val:.4f}")
```

---

5. Cluster Profiling – Average Engagement Scores by Cluster

Task Description
Describe user clusters by calculating average engagement within each cluster.

Why This Matters
Gives behavioral insights into clusters, aiding in strategic targeting and personalization.

How to Do It

1. Group data by `ClusterLabel`.
2. Calculate the mean of `EngagementScore`.
3. Optionally, rename and round for reporting.

Tools/Modules

- `pandas`

Output
Cluster-wise average engagement scores.

python
CopyEdit

```
cluster_profile =
df.groupby('ClusterLabel')['EngagementScore'].mean().reset_index()

# Optional: rename and round for clarity
cluster_profile = cluster_profile.rename(columns={'EngagementScore':
'AvgEngagement'})
cluster_profile['AvgEngagement'] =
cluster_profile['AvgEngagement'].round(2)

print(cluster_profile)
```