UrBox Test Answer

Candidate Name: Ngo Thanh Luan
Position Applied For: Business Intelligence Analyst

Important Notice

Analysis Scope Clarification This report is **a User Lifecycle Analysis**, not a **Loyalty Scoring** analysis. The age groups used in this test (New, Early, Growth, Loyal) are based on UrBox Age —the time between a user's first and last redemption. These labels indicate how long users have been active on the platform, but they do not represent true customer loyalty levels. To evaluate real user loyalty, additional metrics such as frequency, recency, and value would be required (e.g., RFM or engagement scoring). In this analysis, we focus only on behavior differences between users with different platform usage durations.

PART I

SQL Test

Source Code:

Please check file “SQL_Code.sql” in folder “Code”

Result

100 %

Results Messages

	user_id	first_brand_id	second_brand_id	first_brand_redeemed_date	second_brand_redeemed_date	last_brand_id	last_redeemed_date	number_of_brands
1	108465	397	NULL	2025-01-19 18:44:23	NULL	397	2025-01-23 20:05:49	1
2	135042	890	NULL	2025-05-11 10:27:02	NULL	890	2025-05-11 14:41:30	1
3	218587	23	NULL	2025-07-27 20:16:58	NULL	23	2025-07-27 20:16:58	1
4	410610	82	NULL	2025-07-01 17:07:12	NULL	82	2025-07-01 17:07:12	1
5	427212	992	1511	2025-06-21 07:22:19	2025-07-06 21:04:40	1511	2025-07-06 21:04:40	2
6	539864	1511	NULL	2025-05-10 21:36:24	NULL	1511	2025-05-10 21:36:24	1
7	735746	396	NULL	2025-01-09 00:30:39	NULL	396	2025-01-09 00:30:39	1
8	890259	593	NULL	2025-07-16 18:21:55	NULL	593	2025-07-16 18:21:55	1
9	1385395	315	34	2025-07-19 02:05:19	2025-07-23 01:31:23	315	2025-07-23 01:31:23	2
10	1537715	395	NULL	2025-02-05 16:30:12	NULL	395	2025-02-05 16:30:12	1
11	1571771	910	NULL	2025-06-26 21:36:50	NULL	910	2025-07-26 22:23:21	1
12	1648417	82	NULL	2025-07-13 16:37:45	NULL	82	2025-07-13 16:37:45	1
13	1649747	654	NULL	2025-06-16 18:35:33	NULL	654	2025-06-16 18:35:33	1
14	1734287	882	NULL	2025-07-10 03:21:14	NULL	882	2025-07-10 03:21:14	1
15	1743610	673	82	2025-04-11 03:22:38	2025-04-11 03:44:27	673	2025-05-10 17:02:22	2
16	1803669	138	1511	2025-06-28 14:13:43	2025-07-02 03:23:32	1511	2025-07-15 15:34:02	2
17	1830892	552	NULL	2025-05-26 22:22:28	NULL	552	2025-05-26 22:22:28	1
18	2970281	873	NULL	2025-07-25 03:35:43	NULL	873	2025-07-25 03:35:43	1
19	3680103	1511	992	2025-06-07 23:40:10	2025-07-30 01:54:07	1511	2025-07-30 01:54:07	2
20	5054003	82	NULL	2025-03-20 11:09:49	NULL	82	2025-03-20 11:09:49	1
21	5844233	0	82	2025-01-04 15:38:28	2025-07-06 19:51:31	992	2025-07-12 18:06:13	3

Analytics Test 1

I. Analysis Summary (Only for User & Transaction Numbers)

Based on the UrBox dataset, I have aggregated the key numbers for each age group. The results show clear differences in both the number of users and the level of activity across the groups.

Segmentation

New Users (0–7 days)

Early Users (8–30 days)

Growth Users (31–90 days)

Loyal Users (>90 days)

	age_group	user_count	transaction_count	avg_transactions_per_user
0	New User	481	898	1.9
1	Early User	80	416	5.2
2	Growth User	185	1303	7.0
3	Loyal User	254	3035	11.9

1. **New Users (481 users, 898 transactions)**
   - New Users have an average of 1.9 transactions per user, which is the lowest among all groups.
   - This means most new users only redeem once or twice.

2. **Early Users (80 users, 416 transactions)**
   - Early Users have 5.2 transactions per user, which is a strong increase compared to New Users.
   - However, the group size is very small (only 80 users).
   - This suggests that many New Users do not continue into the Early stage, indicating a retention issue.

3. **Growth Users (185 users, 1,303 transactions)**
   - Growth Users make 7.0 transactions per user, showing consistent engagement.
   - This group is stable and shows good potential to become long-term users.

- Their activity level is higher than Early Users and contributes a medium share of total transactions.

4. **Loyal Users (254 users, 3,035 transactions)**
   - Loyal Users have the highest activity with 11.9 transactions per user.
   - This group also produces the largest total number of transactions (over 3,000).
   - They play an important role in the overall performance of the platform.

II. Brand Switching Analysis by Age Group

This table shows how many different brands each user redeems inside each age group. It helps us understand how user behavior changes when they stay longer on UrBox.

	age_group	min_value	max_value	mean_value	median_value	mode_value	count_value
0	New User	1	4	1.10	1.0	1	481
1	Early User	1	8	1.94	2.0	1	80
2	Growth User	1	7	1.98	2.0	1	185
3	Loyal User	1	10	2.87	3.0	2	254

1. **New Users (481 users)**
    - Mean = 1.10 brands, Median = 1, Mode = 1
    - Most New Users redeem only 1 brand.
    - This is normal for a first-stage user, as they are still trying the platform.
    - The range is 1 to 4 brands, but the majority stay at 1 brand.

<u>Meaning</u>:
New Users have very simple behavior and low brand exploration. Our main focus here should be helping them discover a second brand and move to the next stage of the funnel.

2. **Early Users (80 users)**
    - Mean = 1.94 brands, Median = 2
    - Early Users begin to explore more brands compared to New Users.
    - Even though the group is small, their brand switching almost doubles.

<u>Meaning</u>:
This group shows that once users make more than 1–2 transactions, they start exploring new brands. This is the transition stage where recommendations and light incentives can help them expand their usage.

3. **Growth Users (185 users)**
    - Mean = 1.98 brands, Median = 2
    - Brand switching remains similar to Early Users, but this group is larger and more stable.
    - Users redeem 1–7 different brands.

<u>Meaning</u>:
Growth Users maintain steady exploration behavior.
They use more features and brands regularly, but the pattern is still moderate.
This is a good stage to strengthen habits and build long-term engagement.

4. **Loyal Users (254 users)**
    - Mean = 2.87 brands, Median = 3, Mode = 2
    - Loyal Users explore the most brands and have the widest range (1 to 10 brands).
    - They have almost 3× more brand diversity than New Users.

    Meaning:
    This group shows the highest level of trust and activity on the platform.
    They are open to trying many brands and different categories, which makes them suitable for cross-sell and upsell strategies.

Key Insights

Brand exploration increases with user age.
New Users stick to 1 brand, while Loyal Users explore up to 10.

The biggest behavior change happens after the New stage.
Users who stay become more open to trying different brands.

Loyal Users have the highest brand diversity, meaning they interact with multiple brand clusters.

What we should do

New Users: guide them to try a second brand

Early Users: recommend similar or related brands

Growth Users: strengthen habits with category promotions

Loyal Users: apply cross-sell & upsell, promote premium brands

III. Analysis of Brand Cluster Differences Across UrBox Age Groups

This section analyzes whether users in different UrBox Age Groups redeem different sets of brands. To answer this question, we use two pieces of evidence:

- Top 5 most redeemed brands per age group (brand concentration)
- Total number of distinct brands redeemed per age group (brand diversity)

Combining both allows us to understand not only which brands users prefer, but also how broad or narrow their brand clusters are.

1. Summary of Distinct Brand Counts

	age_group	brand_count
0	New User	48
1	Early User	43
2	Growth User	47
3	Loyal User	59

2. Top 5 Brands per Age Group (Ranked by Transactions)

	age_group	brand_id	transaction_count
0	New User	1511	276
1	New User	395	99
2	New User	82	93
3	New User	882	52
4	New User	910	38
5	Early User	82	71
6	Early User	1511	48
7	Early User	882	38
8	Early User	395	36
9	Early User	552	34
10	Growth User	1511	322
11	Growth User	552	160
12	Growth User	82	155
13	Growth User	395	91
14	Growth User	882	72
15	Loyal User	1511	615
16	Loyal User	552	521
17	Loyal User	82	495
18	Loyal User	7	136
19	Loyal User	673	133

3. Combined Analysis (Brand Concentration + Brand Diversity)

3.1 New Users – Wide pool (48 brands) but highly concentrated usage

Although New Users touched 48 different brands, the vast majority of transactions come from only 5 brands (especially 1511). This means New Users behave with broad but shallow brand exploration.

Meaning:
New Users don't truly explore clusters; they mostly stick to 1–2 dominant brands.

3.2 Early Users – Brand diversity drops (43 brands), behavior stabilizes

Early Users redeem fewer distinct brands than New Users (43 vs 48). Their top brands remain almost identical to New Users. This shows that Early Users begin to stabilize preference, focusing on the brands they like rather than trying new ones randomly.

Meaning:
Early Users form clearer brand preferences and narrow their cluster.

3.3 Growth Users – Diversity increases again (47 brands) + higher depth

Growth Users expand diversity from 43 → 47 brands. They show increased usage of brand 552 and maintain strong usage of the known brands. This stage represents intentional exploration, not random like New Users.

Meaning:
Growth Users discover additional clusters and engage deeper across brands.

3.4 Loyal Users – Highest diversity (59 brands) + new categories emerge

Loyal Users redeem the most distinct brands: 59 brands. They introduce new top brands (7 and 673) not seen in earlier stages. This means they explore new categories, not only the popular ones.

Meaning:
Loyal Users are multi-cluster users with the widest and deepest brand behavior.

4. Conclusion

Users in different UrBox Age Groups redeem different sets of brands.

- **New Users** use a few main brands, even though the whole group touches many brands. Their behavior is still simple and not very deep.
- **Early Users** show more stable behavior. They focus on a smaller and clearer set of brands that they like, and do not explore much at this stage.
- **Growth Users** start to open up again. They use more brands and show stronger activity with both old and new brands.
- **Loyal Users** have the widest brand usage. They redeem many different brands, including some brands that do not appear in younger groups.

Overall, as user age increases, brand usage becomes more stable, broader, and more diverse.

IV. BASKET ANALYSIS TO FIND BRAND IDS FOR CROSS-SELL

In this case, I use the Apriori method to find which brand_id we should cross-sell.

First, I need to calculate some basic metrics for Apriori:

1. Support

Support shows how often a **brand** appears in all transactions.

$$\text{Support}(A, B) = \frac{\text{Number of users who redeemed both A and B}}{\text{Total number of users}}$$

2. Confidence

Confidence shows the percentage of times users redeem **brand B** when they redeem **brand A**.

$$\text{Confidence}(A \to B) = \frac{\text{Number of users who redeemed both A and B}}{\text{Number of users who redeemed A}}$$

3. Lift

Lift compares how often **brand A and brand B** appear together versus how often they would appear by chance.

$$\text{Lift}(A \to B) = \frac{\text{Confidence}(A \to B)}{\text{Support}(B)}$$

Here, I calculated the rules using the Apriori library in Python. Let's look at the results.

ASSOCIATION RULES TABLE				
antecedents	**consequents**	**support**	**confidence**	**lift**
(992)	(552)	0.029	0.453125	4.157110
(552)	(992)	0.029	0.266055	4.157110
(7)	(552)	0.025	0.438596	4.023821
(552)	(7)	0.025	0.229358	4.023821
(552)	(673)	0.031	0.284404	3.843293
(673)	(552)	0.031	0.418919	3.843293
(82)	(673)	0.038	0.163793	2.213420
(673)	(82)	0.038	0.513514	2.213420
(82)	(992)	0.027	0.116379	1.818427
(992)	(82)	0.027	0.421875	1.818427
(82)	(7)	0.023	0.099138	1.739262
(7)	(82)	0.023	0.403509	1.739262
(82)	(552)	0.043	0.185345	1.700411
(552)	(82)	0.043	0.394495	1.700411

STRONG RULES TABLE				
antecedents	**consequents**	**support**	**confidence**	**lift**
(992)	(552)	0.029	0.453125	4.157110
(7)	(552)	0.025	0.438596	4.023821
(673)	(552)	0.031	0.418919	3.843293
(673)	(82)	0.038	0.513514	2.213420
(992)	(82)	0.027	0.421875	1.818427
(7)	(82)	0.023	0.403509	1.739262
(552)	(82)	0.043	0.394495	1.700411

Note:

- **Association Rules Table**
    - This table contains all generated association rules, along with their key metrics such as support, confidence, and lift.
    - It provides a complete view of every possible relationship identified between brands.

- **Strong Rules Table**
    - This table is a filtered version of the association rules, showing only the most meaningful and actionable relationships.
      Rules are included in this table only if they meet all three threshold conditions:
    - Support ≥ 0.02 (the rule must appear in at least 2% of users)
    - Confidence ≥ 0.30 (at least 30% of users who used the antecedent also used the consequent)
    - Lift ≥ 1.20 (the relationship is at least 20% stronger than random chance)

<u>These filters ensure that only strong, reliable, and business-relevant rules are selected for cross-sell recommendations.</u>

Based on the Strong Rules criteria (Support ≥ 0.02, Confidence ≥ 0.30, Lift ≥ 1.20), the 7 brand pairs with the strongest cross-sell potential are:

- <u>992 → 552</u>
- <u>7 → 552</u>
- <u>673 → 552</u>
- <u>673 → 82</u>
- <u>992 → 82</u>
- <u>7 → 82</u>
- <u>552 → 82</u>

<u>Summary Meaning</u>

- **552 and 82** are the strongest target brands for cross-sell (they appear in most rules).
- **992, 7, and 673** are good starting brands because they lead users to many other brands.

These insights can help improve product suggestions, create better combos, increase conversion, and build cross-sell flows based on real user behavior.

What we should do after getting these metrics

1) **In-App Cross-Sell Recommendations**
   - Show the target brand right after the user redeems the source brand.
2) **Combo / Bundle Offers**
   - Create special combo discounts for strong brand pairs to increase redemption and order value.
3) **Remarketing & Retargeting**
   - Send push or email messages to remind users to redeem the target brand based on their source brand history.
4) **Loyalty / Progressive Offers**
   - Give a reward for the target brand after the user redeems the source brand to increase LTV.
5) **Co-Marketing Partnership**
   - Recommend strong-related brands to run joint campaigns and drive traffic between them.

<u>**Analytics Test 2**</u>

I. How to Increase the Average Number of Brands Users Engage In

A. Analysis Plan

This plan explains how to analyze user behavior and find ways to increase the number of brands they redeem.

1. **Measure current brand engagement**
   a. Count how many brands each user redeems
   b. Calculate average and median brands per user
   c. Compare results across age groups (New, Early, Growth, Loyal)

   Identify which group has low brand diversity

2. **Identify low-engagement users**
   a. Find users who redeem only 1–2 brands
   b. These users need support or guidance to try more brands
3. **Analyze brand overlap**
   a. Check which brands are often redeemed together
   b. (Brand A → Brand B)

   Find natural cross-sell opportunities

4. **Analyze time behavior**
   a. See when users start trying new brands
   b. Check if New Users stay with only a few brands

   Identify the best timing to recommend new brands

5. **Identify possible barriers**
   a. Low awareness of other brands
   b. Repeated or limited brand suggestions
   c. No incentive to explore new brands
6. **Recommend simple actions**
   a. Better brand recommendations
   b. Cross-sell based on user behavior
   c. Bonus or rewards for trying a new brand
   d. Highlight new brands in the app

B. What Customer Data We Should Collect

To help users try more brands, we should collect the following data:

1) **Brand Data**
    a. Brand category
    b. Brand tags (type, price, voucher style)

    Helps us recommend the right brand.

2) **User Profile**
    a. Age, gender, location
    b. Device type

    Helps us understand different user groups.

3) **User Behavior**
    a. Clicks on brand pages
    b. Search keywords
    c. Time spent on brand pages

    Shows which brands the user is interested in.

4) **Transaction Details**
    a. Order of brands redeemed
    b. Time between redemptions

    Helps identify cross-sell opportunities.

5) **Campaign Interaction**
    a. Promotions seen
    b. Notifications opened
    c. Offers clicked

    Helps understand motivation to try new brands.

6) **User Feedback or Habit**
    a. Ratings or liked brands

    Helps improve personalized suggestions.

II. How To Increase The Average Frequency Of Transactions Per User Each Month

A. Analysis Plan

1. **Measure current transaction frequency**
   a. Count how many transactions each user makes per month
   b. Calculate average and median transactions per user
   c. Compare across age groups (New, Early, Growth, Loyal)

   <u>Identify low-frequency groups</u>

2. **Segment users by activity level**
   a. Users with 0–1 transactions/month
   b. Users with 2–4 transactions/month
   c. High-frequency users

   <u>Helps find which user groups need support</u>

3. **Analyze behavior differences**
   a. Compare low-frequency vs high-frequency users
   b. Check time between transactions
   c. Check which brands high-frequency users prefer

   <u>Helps understand what drives active users</u>

4. **Study drop-off patterns**
   a. When do users stop redeeming?
   b. How long from last transaction to churn?

   <u>Helps find the right time to re-engage them</u>

5. **Identify barriers to higher frequency**
   <u>Possible reasons:</u>
   a. Not enough relevant offers
   b. Users forget to redeem
   c. No reminders or weak notifications
   d. Limited voucher availability
   e. Poor user experience
6. **Propose actions**
   a. Stronger notifications or reminders
   b. Personalized brand suggestions
   c. Monthly reward or loyalty points
   d. “Come back” promotions
   e. Auto-suggest vouchers when balance is available

B. **What Customer Data We Should Collect**

To understand why users do not redeem more often, we should collect:

1. **Transaction Behavior**
    a. Time between transactions
    b. First transaction date
    c. Last transaction date
    d. Monthly spending or voucher usage

    <u>Helps identify churn patterns and user habits.</u>

2. **User Engagement Data**
    a. App open frequency
    b. Daily/weekly active time
    c. Clicks on promotions or banners

    <u>Helps understand how active a user is.</u>

3. **Campaign & Notification Data**
    a. Which notifications users open
    b. Which promotions they view or ignore
    c. Response to discount campaigns

    <u>Helps optimize re-engagement and reminders.</u>

4. **User Profile**
    a. Age, gender, location
    b. User segment (New, Early, Growth, Loyal)

    <u>Helps target the right users with the right offers.</u>

5. **User Preferences / Habit**
    a. Favorite brands
    b. Categories used often
    c. Time-of-day or day-of-week activity

    <u>Helps offer relevant vouchers to increase usage.</u>

PART II

Based on the transaction data and gift information, we evaluated the campaign across **three key objectives**: user acquisition & retention, used ratio, and profitability.

1. User Acquisition & Retention – Good acquisition, weak retention

- The campaign successfully attracted many new users (high number of first-time redemptions).
- However, **most users only redeemed once**, and the repeat-user rate is low.
- This indicates the campaign brings people in, but does not keep them active long term.

Insight:
The client needs stronger engagement flows (reminders, bonus points for second redemption, personalized suggestions).

2. Used Ratio – Strong performance

- The overall **used ratio is high**, meaning most redeemed vouchers are actually used.
- Several categories (e.g., Café/Thức uống, Siêu thị) show very healthy usage.
- A few categories show lower used ratio and longer delays, which may require review.

Insight:
The campaign is effective in getting users to use vouchers instead of abandoning them.

3. Profitability (PnL) – The campaign is currently unprofitable for UrBox

- After applying the **client_discount_rate = 0.02**, UrBox revenue per gift is lower than the cost of buying the gift from merchants.
- Many gifts have very low merchant discounts (0–1%), but UrBox only earns 2% from the client.
- This creates **negative margins on the majority of used vouchers**.
- Profit heatmap and gift-level table highlight several gift IDs with significant losses.

Insight:
The campaign structure does not cover UrBox's cost → **UrBox is losing money per used voucher**.

Final Conclusion

The campaign is successful in attracting users and maintaining a good used ratio.
However, **it is not financially sustainable** because:

- Merchant discounts are too low
- Client discount rate (2%) is not enough to cover cost
- High-volume categories are generating negative profit

To improve profitability, UrBox needs to renegotiate rates, rebalance gift categories, or introduce minimum discount thresholds.

Please check the Power BI file for all metrics used in the conclusion. Thank you.