

Subject: Key Insights on User Growth Trends and Data Quality Issues

Hi team,

Hope you all have a great day! I would like to share some key finds from my analysis and highlight a few data issues which need to be addressed.

Key data quality issues:

1. User data
 - a. ~99.5% of transaction entries are ***missing*** corresponding user information data
 - b. This could potentially indicate a data pipeline or data storage issue or a mismatch in the transaction dataset that has only 88 days
 - c. Gender values should be combined and standardized
2. Product data
 - a. ~45.4% of BARCODE values in the transaction table are ***missing*** in the Product table, so we are missing essential brand and category information, which should be difficult to analyze user purchasing patterns.
 - b. Clean up the CATEGORY_1 column, particularly rows labeled as 'Needs Review'.
3. Outlier data:
 - a. User age outliers, for instance: < 5-year-old or > 100 year old
 - b. Some dates created before APP launch, which could be some testing data entry or error data, should investigate and clean these
 - c. Some transaction records have receipts scan date earlier than purchase data

Insights summary:

1. Our data indicates potential receipts quantity anomaly:
 - a. From the analysis, on average each receipt has ~ 2 items
 - b. Only 278 out of 24440 (~1%) receipts have more than 5 items in each purchase, which is usually low for our TOP 5 stores: WALMART, DOLLAR GENERAL STORE, ALDI, TARGET, KROGER
 - c. Most people will go shop for more items in these stores, but in our database, our avg quantity shopped from this store is ~ 1 item
2. Frequency in grocery shopping
 - a. Our transaction data shows within June 2024 - Sep 2024, on average scanned receipts by user is around 1.4, which is a significantly lower than expected

- b. Most people go grocery shopping 3-4 times a month, but our data shows much lower frequency
 - i. People might forget to scan their receipts
 - ii. Lack of attractive deals and promotions during this period
 - iii. Potential user churn or less engaged with the app
- 3. Fetch user growth is moving slowly
 - a. Data time range: Jan - Aug from 2017 – 2024 due to incomplete data in 2024.

	member_year	previous_year	users	last_year_users	user_diff	yoy_user_growth
0	2015	2014	36	21	15	71.428571
1	2016	2015	59	36	23	63.888889
2	2017	2016	258	59	199	337.288136
3	2018	2017	1109	258	851	329.844961
4	2019	2018	3590	1109	2481	223.715059
5	2020	2019	10103	3590	6513	181.420613
6	2021	2020	12979	10103	2876	28.466792
7	2022	2021	17718	12979	4739	36.512828
8	2023	2022	11335	17718	-6383	-36.025511
9	2024	2023	10652	11335	-683	-6.025584

- b. 2017 - 2018: rapid growth period after APP launch on average 330%+ each year
- c. 2019 - 2020: high growth ~ 200% YoY; strong user acquisitions
- d. 2021 - 2022: slower growth, compared to previous years, might be due to Covid, inflation, potentially changed user behavior
- e. 2023 - 2024: negative growth; could be a result of external economic reasons (inflation, layoffs, low job opening in the market) reduced purchasing power

If we look at post covid after 2020, the growth trend is similar as GDP and opposes to inflation rate trends

Next Steps:

1. Investigate above-mentioned Data Quality issues:
 - a. We need to clarify whether the data issue is due to an error in data loss, entry errors, or data pipeline
 - b. Suggest to standardize the columns: such as: gender
2. To better understand why user growth is slowing down, we can:
 - 1) Review our **marketing efforts**, was there a shift in marketing strategy or budget cut?

- 2) Understand **user retention and loyalty**:
 - a) Analyze reasons why user churn, product driven? Are redemptions less attractive? People concerned about their personal data?
 - b) What are the features to drive user activities on APP? New app feature launch?
- 3) **Competitor activities**:
 - a) Are there competitors gaining market share or brand awareness? For instance: Ibotta
- 4) **External factors**:
 - a) User purchasing behaviors are impacted by these, we should continue monitoring these factors and tailor our app functions.

Please let me know if you'd like to discuss this further or you have any questions or insights would like to share. Happy to explore these further and collaborate on the next steps!

Best regards,

Shuting