

Subject: Data Quality Insights & Fetch's Growth Trends

Hi Team,

I hope you're doing well. As part of my recent analysis of our datasets, I identified key data quality issues that may impact reporting accuracy and decision-making. Additionally, an interesting trend emerged regarding Fetch's **year-over-year (YoY) growth**, which is worth highlighting. Below is a summary of my findings:

Key Data Quality Issues:

1. User to Transaction Mapping Inadequacy

- Only **0.51%** of transaction user IDs are found in the user dataset, limiting our ability to analyze customer behavior.
- **Outstanding question:** Could this be due to missing user records, or are transactions originating from sources not linked to our user database?

2. Barcode Format Inconsistencies

- Barcodes in transactions and product datasets vary in length (7 to 14 digits), making product mapping inconsistent. Additionally, storing barcodes as integers can cause leading zeros to be dropped, leading to mismatches between datasets.
- **Solution:** Convert all barcodes to strings and apply zero-padding to ensure a standardized 14-digit format across datasets.

3. Duplicate Transactions Due to Missing Data

- Some transactions were **duplicated** where either FINAL_SALE was empty or FINAL_QUANTITY was zero, leading to inflated counts.
- **Solution:** Dropped transactions where FINAL_SALE was empty or FINAL_QUANTITY was zero, as they do not represent valid purchases.

4. High Volume of Unknown Brands and Manufacturers

- A large proportion of product records contain 'Unknown' values for brand and manufacturer details, limiting insights into vendor performance.
- **Outstanding question:** Are these due to incomplete vendor submissions, or should we implement data enrichment processes?

5. Misalignment Between Product and Transaction Barcodes

- **41% of transaction barcodes** do not match product barcodes even after standardization, affecting sales tracking.
- **Outstanding question:** Is this due to missing product entries, or do different stores use different barcode mappings for the same item?

6. Timestamp Gaps & Unrealistic Dates

- Some transactions have purchase dates **after** the scan dates (~46 transactions)
- **Outstanding question:** Could this be due to **system errors in timestamp logging**, or are there scenarios where purchases are recorded after scans due to delayed processing?

Interesting Trend: Fetch's YoY Growth & Recent Decline

The **YoY growth analysis** revealed an **explosive growth phase** between **2017-2022**, where Fetch grew by **820% in 2017, 236% in 2018, and 227% in 2019**. However, recent years show a sharp decline:

- **2023: -42.3% growth**
- **2024: -24.8% growth**

This indicates **either a market saturation effect, reduced user acquisition, or shifts in user engagement**. It would be helpful to investigate whether this decline is due to:

- External market trends (e.g., competition, economic changes)
- Internal factors (e.g., changes in rewards, user experience, or marketing strategies)

Next Steps & Questions for Discussion:

To improve data integrity and understand user trends, we may want to:

- **Enhance data integration** between users and transactions to improve mapping accuracy.
- **Standardize barcode formats** and ensure alignment between product and transaction datasets.
- **Investigate recent YoY declines** to understand what factors might be driving lower engagement.

Would love to hear your thoughts on these findings, and if there's someone best suited to discuss potential solutions. Happy to provide additional details if needed!

Best,
Dhanesh Akolu