

Subject: Data Investigation Summary & Key Insights

Hi team,

I've completed an initial analysis of our transaction and product data, identifying key data quality issues, while highlighting trends and uncovering insights. Below is a summary of my findings and some next steps where your input would be valuable.

Key Data Quality Issues & Open Questions

1. Missing & Unmatched Data:

- **17,603 transactions (35%)** reference a **USER_ID** that does not exist in the user table.
- **4,465 transactions (9%)** contain barcodes that do not match any product in the product table.
- Some transactions are missing barcodes, which makes it impossible to categorize them correctly.
- **Outstanding Question:** How are these missing values affecting our ability to analyze user behavior and product performance?

2. Data Formatting Issues:

- **FINAL_QUANTITY** had non-numeric values like "zero", which were corrected.
 - **FINAL_SALE** had some missing values, requiring assumptions for analysis.
 - **Outstanding Question:** Is **FINAL_SALE** always meant to represent the total transaction cost, or do we need to factor in **FINAL_QUANTITY** to get accurate sales figures?
-

Interesting Trend in the Data

Health & Wellness is a key driver of sales, particularly for Millennials.

- **58% of sales for Millennials, Gen X and Baby Boomers** come from Health & Wellness products
- **Gen Z contributes virtually no sales to this category.**

Opportunity: If Fetch wants to engage younger audiences, targeted promotions for Health & Wellness products could be an avenue to explore.

- **Brand Loyalty Among Long-Term Users:** COCA-COLA, ANNIE'S HOMEGROWN GROCERY, and DOVE are the top-performing brands among users who have been with Fetch for at least six months. This suggests strong brand loyalty and provides an opportunity to explore targeted promotions or partnerships.
-

Outstanding Questions and Request for Action:

1. Can we confirm if the lack of Gen Z transactions is a data issue or reflective of actual behavior?
2. Are there known issues with product data completeness, particularly for brand details?
3. It would be helpful to get context on any recent changes in how user or transaction data is collected that could explain these anomalies.

Any insights or help in addressing these questions would be greatly appreciated. Let me know if you need more details or a deeper dive into the findings.

Looking forward to your feedback!

Best regards,
Drew King