

Subject: Data Investigation Summary – Key Issues, Trends, and Next Steps

Hi Product Leader,

I've reviewed the data from our recent analysis, focusing on **data quality, trends, and areas that need further clarification**. Below is a summary of my findings:

Key Data Quality Issues & Outstanding Questions

- **Inconsistent Data Formatting:** Some key fields (e.g., **FINAL_SALE**, **BIRTH_DATE**) have missing or incorrectly formatted values, making certain analyses unreliable.
- **Duplicate Entries:** I observed duplicate **RECEIPT_IDS** in the transaction data, which may impact the accuracy of sales tracking.
- **Unclear Category Mapping:** Some products lack proper categorization, particularly in **CATEGORY_4** that has many missing values, which affects insights into purchasing behavior.
- **Missing Barcodes:** A significant portion of transactions is missing **BARCODE** (~ 12%), limiting our ability to analyze product details accurately.

Interesting Trend in the Data

One notable finding is that amongst the **power users** (top 5% users for receipts scanned and transactions submitted), **Boomers** (ages 60+) account for the **second highest transaction volume but highest total spending**. They primarily shop at **traditional grocery stores** with a **high variety of store choices** and show engagement in a **diverse number of categories**. Top categories include health & wellness, snacks, alcohol, beverages and dairy.

Millennials (28-43 years old) have the **lowest transaction volume but the second highest total spending**. They have a **lower variety of store choices**, indicating different engagement patterns, that spread amongst **three main categories**, health & wellness, snacks and pantry. They are a group with the **highest spend on snacks** and the **only group to show purchases in the pantry** category amongst the power users.

Request for Action

To improve the accuracy and depth of our insights, we need:

1. **Clarification on Data Cleaning Procedures:** Do we have a standard process for handling missing or duplicate transactions?
2. **Better Product Categorization:** Can we get a verified mapping of **CATEGORY_4** for products missing this information?

3. **Barcode Data Completeness:** Are there known issues with how barcodes are recorded, and can we retrieve that missing detail?
4. **Feedback from Business Teams:** Would it be valuable to segment marketing efforts based on **generation-driven** shopping patterns? Fetch's most valuable segment consists of Boomer and Millennials users, who scan frequently and spend more. Personalized exclusive rewards or promotions targeting Millennials (loyalty programs, bulk discounts) and Boomer (health & wellness incentives) could increase retention and spending. Store-specific deals based on generation trends could optimize Fetch's rewards program as well.

Let me know how you'd like to proceed or if you need additional details. Happy to discuss further!

Thanks,
Mei