

Subject: Fetch Data Insights: Quality Review, Key Trends, and Next Steps

Dear Fetch,

I’ve completed the preliminary analysis of Fetch’s dataset, focusing on data quality, trends, and potential action points. Below is a summary of key findings and outstanding issues that need addressing.

Key Data Quality Issues and Questions:

1. Transactions Table:

- **Missing Data:** 5,762 rows lack BARCODE values, and 12,765 rows have null FINAL_SALE entries.
- **Inconsistent Entries:** FINAL_QUANTITY and FINAL_SALE contain zero values in some rows, with identical entries that have correct values.
- **Duplicates:** 335 rows appear duplicated—are these duplicate transactions, or should they be treated as repeat purchased items?
- **Anomalous Dates:** 47 receipts show a SCAN_DATE earlier than PURCHASE_DATE. How should these be interpreted? Could they be manual errors or potential fraud cases?
- **Invalid Barcode Values:** BARCODE values of -1 exist, which appear erroneous.

2. Users Table:

- **Incomplete Data:** Roughly 3.7% of BIRTH_DATE and 5.9% of GENDER entries are missing. Is there a preferred method for handling these gaps?

3. Products Table:

- **Missing Data:** Over 4,000 items lack a BARCODE, and CATEGORY_4 has 92% missing values. Is this field necessary for our analysis?
- **Duplicate Barcode:** 2 brands have the same BARCODE. Which is the correct one?

	CATEGORY_1	CATEGORY_2	CATEGORY_3	CATEGORY_4	MANUFACTURER	BRAND	BARCODE
28421	Health & Wellness	Hair Care	Hair Color	NaN	HENKEL	SCHWARZKOPF	052336919068
213340	Health & Wellness	Hair Care	Hair Color	NaN	HENKEL	SCHWARZKOPF	017000329260
304021	Health & Wellness	Hair Care	Hair Color	NaN	HENKEL	GÖT2B	017000329260
709607	Health & Wellness	Hair Care	Hair Color	NaN	HENKEL	GÖT2B	052336919068

Interesting Trend in the Data

Brand Loyalty Among Long-Term Users: Analysis shows that among users with accounts older than six months, the top 5 brands by sales were CVS, Dove, Trident, Coors Light, and Tresemmé. These brands span personal care, snacks, and beverages, showing loyalty among long-standing Fetch users.

Implications:

- **Long-Term Loyalty:** Users who stay on the platform for over six months seem to develop consistent purchasing habits with specific brands. This loyalty could be leveraged by Fetch to reward users who repeatedly purchase from these brands.

- **Retention Strategies:** Fetch could implement loyalty programs or long-term incentives focused on these brands, rewarding users for sustained engagement. These promotions could help maintain retention and prevent churn.
- **User Segmentation:** This data can also help refine user segmentation strategies. Understanding the buying behavior of long-term users with specific brands enables Fetch to customize its engagement efforts further, catering to the preferences of its most loyal customers.

Request for Action

To improve the clarity and utility of this data, the following additional information or actions are needed:

- **Clarification on Duplicate Transactions:** Are the duplicated rows actual repeated transactions or data errors?
- **Standardization of Missing Birth Dates:** Should we use a default date (e.g., 1970-01-01) to handle missing birth dates in the Users table?
- **Guidance on Invalid Barcode Handling:** How should we treat the -1 entries in BARCODE? Do they have any special meaning, or should they be removed?
- **Relevance of High-Missing Fields:** Is the CATEGORY_4 field essential despite its high level of missing data?
- **Duplicate Barcode:** 2 brands have the same BARCODE in the products table. Which one should be considered?
- How should we handle rows where FINAL_QUANTITY or FINAL_SALE contain zero values alongside identical entries with correct values? Are these items returned or free products or

Thank you for your time. Addressing these questions would enable more refined insights and reliable findings. Please let me know if any other aspects need adjustment or further exploration.

Best regards,
Sai Pranay Reddy Aare.