

Email_Communication

Email Subject: Data Quality Insights and Key Findings from Fetch Data Analysis

Hi [Recipient's Name]/Business Team/All,

According to the provided data files, I have completed the initial analysis of Fetch's data, focusing on data quality issues, key trends, and insights from transactions, users, and product records. Below is a summary of the findings and recommended actions.

There are some key data quality issues that I have found are -

1. Foreign Key Integrity Issues:
 - Only **91 users** from the **Users table** appear in the **Transactions table**, which seems unusually low for a dataset of 100,000 users.
 - Only **6,562 transactions** out of **50,000** have barcodes matching the **Products table**, indicating a large gap between product and transaction data.
 - If foreign key constraints were enforced, over **18,186 transactions would be removed**, indicating **possible missing or incorrect data linkages**.
2. Duplicate Records & Inconsistencies:
 - **4,025 products** had missing barcodes, which were removed for integrity.
 - **185 products** had duplicate barcodes, leading to data redundancy.
 - Several duplicate transactions were identified and removed.
 - Some transactions had non-numeric values in the quantity field (e.g., "zero")—these were cleaned for accurate analysis.
3. Standardization Issues:
 - The **gender** and **language** fields in the Users table contain inconsistent or unclear values. While this doesn't impact our current analysis, it could pose challenges in future segmentation.

I would like to present some interesting trends that can be looked at -

- One of the key insights from the analysis is that **Dove, Coca-Cola, and Nerds Candy are among the most frequently scanned brands by users aged 21 and over.** This suggests high engagement with these brands and could be useful for targeted promotions or partnerships.
- More, within the **Health & Wellness category**, **Baby Boomers contribute the highest percentage of sales (34.35%),** indicating that this generation is a key demographic for wellness-related products

For further analysis, I would need some clarification on the data gap -

- Can we confirm if the Users table represents all registered users or only a subset of active users because as I have mentioned above not all users are listed from the transaction table into the main user data?
- Why do such a small percentage of transactions match products? Are we missing a link between transactions and product records?
- Are there any known data integrity issues causing these mismatches?

Next Steps

- Standardize fields like **gender** and **language** for better segmentation.
- Review why so many transactions reference products that aren't listed in the **Products table**.
- Investigate whether the **missing users in the Transactions table** were inactive or if there's a data mapping issue.

I'd appreciate any insights from the business team regarding the questions above. Let me know how we can collaborate to improve data consistency, and that will lead to the most accurate conclusions.

Looking forward to your thoughts!

Thanks and Regards
Ayushi Patel