

Third: Communicate with stakeholders

For this question, I am writing to John, my (fictional) Team Lead, informing them about the data quality issues I found.

Subject: Data quality issues with Users, Transactions, and Products datasets

(Body of the mail):

Hello John,

After conducting a thorough Exploratory Analysis for the Users, Transactions, and Products, I came across the following data quality issues that I believe are important for you to know.

- Some BARCODE values in the transaction's dataset are missing, making it hard to link purchases to products.
- The products dataset has placeholder values like "PLACEHOLDER MANUFACTURER" in the MANUFACTURER and BRAND fields.
- Dates (PURCHASE_DATE, BIRTH_DATE) were stored as text, causing issues with time-based calculations.
- The FINAL_QUANTITY field had text values like "zero", which required cleanup for accurate analysis.

Some of the Interesting trends I came across are:

- Coors Light and Dove are leading brands, with total sales of **\$17.48** and **\$11.97**, respectively.
- The leader in this category is **Marketside**, with total sales of **\$4.94**.

These findings suggest strong performance in beverages and personal care, as well as niche success in food-related categories like dips and salsa.

To improve the analysis further, I need:

1. **Clarification on missing FINAL_SALE values**, are these unpaid transactions, discounts, or data entry errors?
2. **Guidance on placeholder entries in MANUFACTURER and BRAND**, can these be mapped to actual values, or should they be excluded?
3. **Confirmation on how to treat inactive users**, should they be excluded from long-term trend analysis?

Let me know if you'd like to discuss these points further or if there's anything else I should focus on.

Thanks,
Akhil