### Hi Kayla Donnell

I wanted to share a quick update on the data investigation I conducted to support our analytics initiatives. While reviewing the datasets related to products, transactions, and users, I discovered several data quality issues and identified one development that may be of interest to the business.

**Key Data Quality Issues Identified:**

**Missing Values: Several critical fields have missing data, anomalies:**

1. In the Product dataset, fields like CATEGORY\_1 to CATEGORY\_4, MANUFACTURER, BRAND, and BARCODE are incomplete.
2. In the Transaction dataset, BARCODE is occasionally missing, which makes it hard to link transactions to specific products.
3. In the User dataset, I found missing entries in BIRTH\_DATE, STATE, LANGUAGE, and GENDER.

**Invalid and Fraudulent Entries:**

1. Some date fields (BIRTH\_DATE and CREATED\_DATE) failed to analyze correctly, showing up as invalid or fraudulent timestamps.
2. I also noticed unusual values in numeric fields like FINAL\_SALE and FINAL\_QUANTITY (e.g., possible zero or negative values), which needs validation.
3. **Duplicate Records:** Some repeated entries across datasets suggest inconsistencies or errors during data collection.
4. **Unclear Fields**: Certain fields such as CATEGORY\_1–CATEGORY\_4 and SCAN\_DATE vs PURCHASE\_DATE are not clearly defined, which limits analyses.

**Interesting Trend Observed:**

One important trend is that private label products (store-owned brands) show a relatively high frequency of transactions compared to national brands in some categories. This suggest strong customer loyalty for pricing advantages that are worth exploring further for strategic promotions based on power users.

**Request for Action:**

**To help us move forward confidently, we need the following:**

1. Clarification or documentation on unclear fields like CATEGORY\_1–CATEGORY\_4, SCAN\_DATE, and FINAL\_SALE. Understanding these fields will ensure accurate segmentation and trend analysis.
2. Business input on how important it is to retain missing values, especially for demographics and product identifiers.
3. Access to metadata if available it would streamline future analysis and reduce assumptions during modeling.

Let me know if you or someone from the data governance and product operations team would be the available to contact for these clarifications. Happy to jump on a quick call and discuss more context if needed!

Thanks,  
Konda Reddy Padala