Hi Team,

I’ve completed an initial analysis of the transaction, user, and product data. Here are some key findings, along with a few outstanding questions that require further clarification.

**Key Data Quality Issues:**

1. Missing Data
   * In PRODUCTS\_TAKEHOME Over 226,000 records are missing manufacturer and brand details, which limits ability to analyze brand-level sales.
   * User Demographics: About 30% of users are missing language data, which affects segmentation analysis.
2. Inconsistent Values
   * Some fields, like FINAL\_QUANTITY, contain non-numeric values such as "zero", which may need to be standardized.
   * BARCODE values are missing in 5,700+ transactions, making it difficult to link products to sales.
3. Duplicate Entries
   * We found 215 duplicate records in the products dataset and 171 in transactions, which may impact accuracy in reporting.

**Interesting Trend in the Data:**

* Health & Wellness products are most popular among Millennials (25-40 years old), contributing to 38% of sales in this category. This suggests that targeted promotions for this demographic could drive further engagement.

**Request for Action:**

To proceed with refining the analysis, we need:

1. Clarification on missing manufacturer and brand details – Should we infer missing data from other sources, or is there a standard process for handling this?
2. Guidance on non-numeric values in FINAL\_QUANTITY – Should "zero" be treated as 0, or does it indicate a different business rule?
3. Feedback on data deduplication – Should we remove duplicate product entries, or do they serve a specific purpose?