**Subject:** Data Quality Review – Key Findings & Next Steps

Hi [Business Leader's Name],

We've completed our initial data analysis and identified some key data quality issues, along with an interesting trend worth further investigation. Below is a summary of our findings, along with a few outstanding questions where your input would be valuable.

**Key Data Quality Issues:**

**Missing or Incomplete Data**

* Final Sale is missing in 12,500 transactions – Are these due to refunds, unprocessed transactions, or a known issue?
* User Data Gaps – Thousands of records are missing key details:
  + Birth Date (~3,675 missing)
  + State (~4,812 missing)
  + Language (~30,508 missing)
  + Gender (~5,892 missing)
* Product Data Gaps – Many product details are missing:
  + Category 1 (~111 missing)
  + Category 2 (~1,424 missing)
  + Category 3 (~60,566 missing)
  + Category 4 (~778,093 missing)
  + Manufacturer (~226,474 missing)
  + Brand (~226,472 missing)
  + Barcode (~4,025 missing)

**Duplicate & Inconsistent Records**

* 48 products have the same barcode but different categories, brands, or manufacturers.
* 25,560 duplicate transactions exist where the Final Quantity or Final Sale values differ.

**Date Anomalies**

* Some transactions have Scan Dates occurring before Purchase Dates. Could this be a data entry issue?

**Outliers in Transaction Data**

* Final Quantity: One extreme value exceeding 250, which may indicate a data entry error or a bulk purchase.
* Final Sale: Three values are between $200-$300, and one extreme outlier is above $400. Should these be included in our analysis, or are they data errors?

**Interesting Trend in the Data:**

**Health & Wellness Products Are More Popular Among Millennials**

* Compared to other generations, Millennials account for the highest percentage of Health & Wellness product sales.
* This could indicate a marketing opportunity or a shift in spending habits. Should we investigate further to refine targeting strategies?

**Outstanding Questions & Action Needed:**

1. Can you confirm if the missing Final Sale values (12,500 transactions) are expected, and if they should be included in our analysis?
2. Are the missing user and product IDs a known issue, or should we investigate potential data ingestion errors?
3. How should we handle duplicate product records with different brand or category info, should we merge them based on barcode or treat them separately?
4. Any insights on why Walmart transactions have lower average sales despite high scan volume?
5. Should we treat the missing product details (e.g., Manufacturer, Brand, Categories) as true gaps, or are they intentionally left blank in some cases?
6. How should we handle outliers in Final Sale and Final Quantity? Are these valid transactions or potential data errors?
7. Should we explore the rise of Health & Wellness purchases among Millennials further, and are there business actions we can take based on this trend?

We’d appreciate your guidance on these issues so we can refine the data and ensure accurate reporting. Let us know if a quick sync would be helpful.

Thanks and Regards,  
Jerome Paul S. Alcantara  
Senior Data Analyst