**Subject: Data Quality Insights & Key Trends – Request for Clarification**

Hi Business Leader's Name,

I’ve completed an initial analysis of our data and wanted to share some key findings, as well as some outstanding questions that require clarification.

**1. Key Data Quality Issues:**

Missing Values: Several critical fields, including CATEGORY\_4 (94% missing), MANUFACTURER (27% missing), and BRAND (27% missing), have significant gaps, which may impact our ability to analyze trends accurately.

Inconsistent Formatting: BIRTH\_DATE and CREATED\_DATE contain timestamps (YYYY-MM-DD HH:MM: SS.000 Z), which might require standardization for consistency in date-based analysis.

Potential Placeholder Values: Some records in MANUFACTURER contain generic labels like "PLACEHOLDER MANUFACTURER," which could indicate missing or unreliable data.

**2. Interesting Trend:**

A preliminary look at purchasing behavior suggests that users aged 21 and older are more likely to scan receipts for premium brands compared to younger users, who tend to engage more with promotional or discounted items. This could have implications for marketing strategies and loyalty programs.

**3. Outstanding Questions & Next Steps:**

a. Source of Missing Data: Do we have insights into why CATEGORY\_4 and BRAND fields are incomplete? Are there upstream data collection issues?

b. Clarification on PLACEHOLDER MANUFACTURER: Should these values be excluded from reporting, or do they require manual categorization?

c. Data Definitions: Are there specific business rules or transformations applied to LANGUAGE codes (e.g., es-419 for Latin American Spanish) that should be considered?

Would love your input on these points so we can refine our approach and ensure our insights are as accurate as possible. Let me know if you’d like to discuss this further.

Thanks,

Zarak Shah