

Fetch Data Analyst Assessment

- Arjun Kumaran

Subject: Data Quality Review and Key Insights

Hello,

I've been analyzing our latest data on products, transactions, and users and wanted to share a summary of my findings, key trends, and a few questions to ensure we're maximizing the potential of our rewards program.

Key Data Quality Issues:

- 1) **Products Data:** Significant portions of data are missing or inconsistent, particularly in the hierarchical category columns. Notably, 92% of data in category4 and 226,474 rows in manufacturer are null, with some placeholder values potentially skewing analysis.

Question: Should placeholder strings be considered missing? Clarifying this will improve our accuracy in understanding product completeness.

- 2) **User Data:** Around 30% of the language column is null, limiting insights on language-based user demographics. There were also some inconsistent age records.
- 3) **Transaction Data:** While generally more complete, this table has integrity issues, with numerous user IDs that don't match records in the User table. Only about 85 unique user IDs align across tables, which could impact the accuracy of user behavior analysis.

Question: Are these discrepancies expected due to how data is gathered, or do they indicate potential gaps in data syncing?

Key Trend:

One interesting insight I found is that Walmart and Dollar General are the two most popular stores among our users. This points to opportunities for tailored rewards at these locations, such as increased incentives or exclusive promotions to drive more engagement. Additionally, Snacks and Health & Wellness are the top-performing product categories. These categories look promising for targeted rewards campaigns and could really boost engagement.

Request for Assistance:

To fully address the data quality issues, it would be helpful to:

- 1) Confirm our approach to handling placeholder strings and high-null columns in the Products data.
- 2) Understand any expected limitations in matching user IDs across tables to refine our reporting assumptions.

I'd appreciate any additional guidance on these points and can discuss further if needed. Thank you, and I'm looking forward to refining our insights.

Best Regards,

Arjun Kumaran