

**Subject:** Data Quality Analysis Findings - Fetch App Receipt Data

Hi Alex K.,

I've been analyzing our Fetch App dataset to uncover key insights and identify areas that need attention. Below is a summary of some key findings and areas where I need additional clarity.

**Key Data Quality Issues**

**1. Duplicate Data:**

- **Transactions:** Each receipt appears 2–12 times, often with inconsistent quantities and sales amounts.
- **Products:** 185 duplicate products (0.02%) have conflicting brand/manufacturer information.

**2. Critical Data Gaps:**

- **User Demographics:** 31% missing language preferences, 12% missing state information
- **Transaction Data:** 12% of transactions can't be linked to products due to missing barcodes

**3. Inconsistent Fields:**

- **Gender Data:** Multiple variations for same gender, for example - "Non-Binary" and "non\_binary".
- **Dates:** For 0.18% of transactions, the "Scan Date" appears before the "Purchase Date"

**Interesting Trend**

I noticed a significant shift in user engagement patterns in recent months:

- July 2024 showed remarkable growth with an 86% increase in active monthly users
- However, this was followed by a concerning decline, culminating in a 72% drop in September 2024

Understanding what drove the surge and subsequent decline could help refine our engagement strategies.

**Request for Action**

To move this analysis forward, I'd appreciate your help with the following:

**1. Data Collection Process Clarity:**

- Could you share any existing documentation about the intended data collection process, particularly around duplicate transactions and products?
- If no documentation exists, would it be possible to connect me with someone from the engineering team who designed these systems?

**2. Business Context:**

- Access to marketing campaign calendars from July-September 2024 to correlate with our user engagement patterns
- Any product release notes or app changes during this period
- Could you confirm if receipts with scan dates earlier than purchase dates are valid entries, or are these likely data entry errors?

**3. Next Steps:**

- I'd like to schedule a 30-minute meeting with you this week to discuss these findings
- Based on your availability, I can propose some time slots for tomorrow or Friday

Once I have this information, I can provide:

- Recommendations for data quality improvements
- A deeper analysis of user engagement patterns
- Potential early warning indicators for future engagement drops

Please let me know which times work best for you to discuss these items.

Best regards,  
Abdeali Arsiwala