

**Subject:** Update on Fetch Rewards Data Analysis & Request for Clarification

Greetings,

I hope you are doing well. My name is Likhita, and I am reaching out to provide an update on the progress of our data analysis for Fetch Rewards, highlight key data quality issues, and seek clarification on a few outstanding questions.

**Summary of Work Completed:**

I have been working with the **Users, Transactions, and Products** tables provided as unstructured CSV files. My approach involved:

1. **Data Cleaning & Transformation** – Handling missing values, standardizing formats, and resolving inconsistencies, deleting duplicates using advanced Python libraries.
2. **Exploratory Analysis** – Solving few business questions and deriving insights using SQL

**Key Data Quality Issues Identified:**

- **Missing Data & Inconsistencies** across all three tables:
  - **Users** – Birth date (~4%), State (~5%), Language (~30%), Gender (~6%)
  - **Transactions** – Barcode (~12%), Final Sale (~25%)
  - **Products** – Category, Manufacturer, Brand, and Barcode fields with missing or placeholder values
- **Data Type Issues:**
  - **Users** – Created Date and Birth Date stored as object types instead of DateTime
  - **Transactions** – Barcode, Scan Date, and Purchase Date stored incorrectly
  - **Products** – Barcode stored as a float instead of a string
- **Inconsistent Values:**
  - **Users** – Variations in Gender values due to free-text entry
  - **Transactions** – Final Quantity contains decimal values, suggesting it might represent weight rather than item count
  - **Products** – Instances where the **same barcode is associated with different manufacturers**, raising questions about product identification

**Interesting Trends:**

**Sales Trend by Purchase Day**

- **Insight:** Sales decline as the purchase day approaches (from Weekend to Weekday). This suggests customers are more likely to make larger purchases earlier during the weekend.

- **Action:** Implement early-to-mid week promotional offers to sustain sales throughout the week.

For example, we can offer Thirsty Thursday offers while scanning receipts of partnered drinks etc.

### Sales Trend by Month

- **Insight:** Sales peaked in July and August but saw a sharp decline in September.
- **Action:** Investigate seasonal trends and customer behavior. Consider running back-to-school promotions or fall discounts to mitigate the drop in September sales.

### User Scanning Behavior

- **Insight:** Most users scan their receipts within the first 1-2 days, with a sharp drop-off after that.
- **Action:** Encourage quicker scanning by offering limited-time rewards or reminders for receipts scanned within 24 hours.

### Year-over-Year (YoY) Growth Trend

- **Insight:** The user base saw massive growth (820%) at its peak in 2017 but has been declining since then.
- **Action:** Conduct user retention analysis (for this we can add a user active flag to know if the user is currently active or has been inactive for a while and rejoined etc. and conduct engagement campaigns).

### Outstanding Questions & Request for Action:

To ensure accurate analysis, I would appreciate further clarification on the following:

- **Final Sale Discrepancy:** There are instances where the **Final Sale** value is zero while the quantity is non-zero. Could you confirm whether this is due to discounts, scanning errors, or another business logic?

**Assumption:** I considered it to be due to discounts and so did not eliminate such values for analysis

- **Final Quantity Interpretation:** Given that Final Quantity contains float values, should we interpret it as **product weight** rather than the number of items purchased?

**Assumption:** I considered it to be product weight since it has float values in it for our analysis

- **Barcode-Multiple Manufacturer Relationship:** In cases where **the same barcode is linked to multiple manufacturers**, is this an expected business practice, or should barcodes be unique to a single manufacturer?

**Assumption:** I considered there might be other datasets or business reasons for not providing unique barcodes for products for my analysis.

These are my initial thoughts on the data. Thank you for your attention and continued support. Your insight on these questions would be valuable in refining the analysis. Please let me know if a meeting would be helpful to discuss further.

Looking forward to your response.

Best regards,  
**Likhita**