Subject: Data Quality Issues Identified & Next Steps for Resolution

Dear [Stakeholder's Name],

I hope you're doing well. As part of our data evaluation, we have identified several **data quality issues** in the Fetch Rewards datasets that may impact reporting and analytics accuracy. Below is a summary of our findings, questions, and proposed next steps.

1. Missing Values
- Users Table:
    - signUpSource (**48 missing**) – Could impact tracking of user acquisition sources.
    - state (**56 missing**) – Limits location-based analysis.
    - lastLogin (**62 missing**) – Could affect user engagement tracking.
- Receipts Table:
    - pointsEarned (**510 missing**) – Affects reward tracking.
    - totalSpent (**435 missing**) – Could cause inconsistencies in purchase value calculations.
    - purchaseDate (**448 missing**) – Makes time-based analysis challenging.
    - finishedDate (**551 missing**) & pointsAwardedDate (**582 missing**) – Could indicate processing delays.
    - rewardsReceiptItemList (**440 missing**) – Missing item-level details on receipts.
- Brands Table:
    - category (**155 missing**) & categoryCode (**650 missing**) – Limits categorization of purchases.
    - brandCode (**234 missing**) – Affects brand-specific analytics.
    - topBrand (**612 missing**) – Missing indicator for key brands.

2. Duplicate Records
- Users Table: **283 duplicate user records**, which may lead to inflated user counts.
- Receipts & Brands Tables: No duplicates found.

3. Invalid Transactions
- **435 receipts have missing or invalid totalSpent values**, potentially affecting revenue calculations.

4. Future Purchase Dates
- No receipts have purchase dates in the future, which is good.

5. Orphaned Foreign Keys
- **148 receipts reference non-existent users**, which may indicate user data deletion or ingestion issues.
- **1,150 brands have barcodes that are not found in any receipt data**, suggesting unused or mismatched product records.

About Performance & Scaling Considerations, as the dataset scales, indexing and partitioning strategies will be essential for optimizing query performance. Implementing data validation rules at ingestion could prevent issues like missing totalSpent values or orphaned records.

Would love to get your input on these issues and discuss the best course of action. Please let me know if you'd like to schedule a quick meeting to go over these findings in more detail.

Looking forward to your thoughts!

Best regards