**Step 4**

**Subject:** Data Quality Issues and Next Steps for Users, Receipts, and Brands Datasets

Hi [Stakeholder’s name],

I hope you are well. I am writing to inform you about several data quality issues identified in our users, receipts, and brands datasets, along with our proposed next steps.

**Users Dataset:**

We found a lot of missing values in the attribute like ‘lastlogin’, ‘signUpSource’, ‘state’. Also, ‘createdDate’ and ‘lastLogin’ should be correctly formatted as datetime instead of strings.

**Brands Dataset:**

We’ve found significant missing values in ‘brandCode’, ‘category’, ‘categoryCode’, ‘topBrand’. Also, ‘barCode’ is stored as integer format but it should be in string format.

**Receipts Dataset:**

Significant gaps in the key attribute columns like ‘brandPointsEarned’, ‘pointsEarned’, ‘barCode’. Also, detected outliers in columns like ‘pointsEarned’, ‘totalSpent’, ‘finalPrice’, ‘itemPrice’.

**Actions Needed**

* Acceptable ranges for ‘totalSpent’, ‘pointsEarned’ columns (As found for example totalSpent for a transaction has a max value around $4500)
* Rules for the data formatting for the fields like data and categorical
* Giving us the business rules to handle missing data.

To ensure efficient handling of the data volumes, we will

* Utilize scalable dataflow tools with the help of cloud infrastructure.
* Optimizing queries and ingestions with indexing and partitioning wherever necessary.
* Develop automated cleaning and validation process

I request you to provide any additional information or guidance you may have. Please contact me back if you have any questions.

With regards,

[NAME and CONTACT]