

Hello [Stakeholder's Name],

Hope you're doing well!

I am writing to discuss some important aspects related to our current data assets and the steps we need to take to ensure their quality and performance. Your feedback is crucial for us to move forward effectively with our analysis. Below are the initial questions and topics I would like to cover:

**1. Data Quality Issues**

- **Document attached** - PFA a document highlighting the data quality issues we encountered during our exploratory analysis on the data.
- **Nulls** - There are null values in the fields of the datasets. What is the nature of these null values? Do these missing values represent data that was not collected or data that is not applicable?
  - i. If it represents data not collected, what are your thoughts on imputing them with a specific value or exclusion from the analysis?
- **Duplicate entries** - Is this a data entry error? How would you recommend we handle this?
  - i. There are duplicate records with same values in the user dataset
  - ii. There are multiple rows for the same brandCode in the brand dataset.

**2. Understanding of the fields in data**

- Could you please let us know when you have time to connect and provide some context over the key fields of the dataset?
- Example - we see there is a field called barcode in the brand dataset and same field in the rewards item list of the receipt dataset. We see that the values do not exactly tie so would like to know how these two fields differ?

Your expertise and detailed feedback on these points will greatly aid us in ensuring our data assets are robust, reliable, and scalable. Please share your thoughts and any additional concerns or suggestions you might have.

Thank you for your attention to these matters.

Regards,

Raghav Goel