

Fourth: Communicate with Stakeholders

Subject: Request for Additional Insights and Guidance on Data Quality Issues

Hi Alexa (fictional name),

I hope this message finds you well. I wanted to bring to your attention some data quality issues I've encountered while working on the fetch rewards data specifically on users, receipts, and brands data and seek your guidance on how to proceed effectively.

During my routine analysis, I noticed some inconsistencies and gaps in the data that could potentially affect the accuracy of our insights and decision-making processes. For example, there are missing values and discrepancies in brands data as well as in users data, which came to light through our standard data validation checks.

To address and resolve these issues efficiently, I have a few questions and requests:

1. Could you provide more context on where our data is sourced from? Knowing more about the origin could help us pinpoint potential causes of these discrepancies.
2. For the columns 'pointsEarned', 'purchasedItemCount', and 'totalSpent', there appears to be a substantial quantity of entries that are inconsistent or abnormal. It would be prudent to review these entries thoroughly to ascertain whether they are valid or if they indicate potential errors in data collection or processing.
3. The users data file contains numerous duplicate records. I recommend a thorough review of our database to remove these excess entries and to implement measures that prevent such issues from recurring.
4. I've noticed that the date formats in our database are not uniform, deviating from the standard MM/DD/YYYY format or similar conventions. It would be advisable to examine our database to confirm that all date fields are captured and maintained consistently.
5. Much of the data, including elements like 'date scanned', 'cpg', 'rewardsReceiptsItemList', is contained within nested dictionaries, necessitating additional data manipulation steps. Eliminating this structure could decrease potential

errors. However, I would appreciate insights into whether this format serves a specific purpose in our product's implementation.

6. Numerous receipts within the dataset featured brand codes that were absent in the brands data. This discrepancy could pose challenges during brand-based aggregation for analytical purposes.

I believe it would be beneficial to discuss these points in detail to ensure clarity and alignment on our next steps. Please let me know when you are available to schedule a meeting, or feel free to share your preferred times, and I will do my best to accommodate. Looking forward to your response.

Best,

Rishitha