

Subject: Queries regarding the Receipts, Brands and Users data

Hi everyone,

After conducting a thorough analysis of the Receipts, Users and Brands data we have a few observations that we wanted to highlight to you:

1. A considerable amount of missing data – We noticed that there is a high number of receipts with missing details such as finishedDate, purchasedItemCount, totalSpentAmount, pointsEarned etc. Could you please shed some light about where these details are being derived from such that any potential issues can be looked at.
2. We observed that the barcode in the Brands table does not match for most of the records with the barcode derived from the rewardsReceiptItemList. Could you provide more details regarding what each of them represent, and if it is expected to be different?
3. Further, we also observed that the Receipts table contains a lot of users (userIDs) which are absent in the Users table. We wanted to understand if there is a different source table for the unmatched users, or if the issue is with missing data in the Users table. This could be an issue because there is a considerable amount of spend on these users which we are not able to map to any user in the Users table.
4. There are a large number of attributes nested within the rewardsReceiptItemList field, we wanted some clarity on the significance of the attributes and if all the attributes are relevant to business
5. In the brands table, we noticed that there are a lot of missing brandCodes which are currently blank in the table.
6. Additionally, we also would like to know if there is a reliable identifier in the Receipts table, which can be mapped to the Brands table to form a relationship between the two. Since there are a lot of NULL brandCodes and barcodes in the Receipts table, in the current state, looking up any data might be misleading due to the presence of null values.

Apart from these issues, we also request you to let us know how frequently the receipts table is expected to be updated, and if it would be okay to archive older receipts data as the table keeps getting refreshed with new data, so that we can limit the size and keep the pipeline fast and efficient.

We would be happy to get on a call with you to discuss these points further in detail

Best Regards,
Akarsh Sonni