Hi XXX,

How are you? I need your help to better understand the transactional data of each customer who scans a receipt. I am working on building a revised structured data model from the unstructured .json files.

I have listed my questions below. Kindly have a look at it and please let me know, if the data model that I have built needs any modifications.

**Questions:**

1. User table has **duplicate** user\_id. Is this allowed?
2. Columns like pointsEarned, purchasedItemCount, totalSpent have **extreme outlier** values. Are these due to error or are they special cases?
3. There exist huge percentage of **nulls** in almost all the tables. Is there a business reason behind the existence of nulls? Can they be removed or imputed? Can any such columns be removed from the data model?
4. Can we receive the item table directly as a separate .json file from the souce? This would reduce the **pre-processing** steps in our end.
5. How should I **link the item table and brand table**? The value of brandcode columns does not match between them. Similarly, value of barcode column doesn’t match between them. Should I link them using oid column and rewardsProductPartnerId column? The values of the above two columns also doesn’t match perfectly.

I have also attached the data model along with this mail for your reference. Feel free to give a detailed explanation to the above questions over mail or over call. Looking forward to hearing from you.

Thank You.

Best,

YYYY