Hi,

I hope this message finds you well. I am currently designing and constructing a database. After going through the data, I found some issues from it. I’m reaching out today to seek some clarifications.

1. Our user data contains many duplicate values. There are 494 records in total in our user data, but there are 283 duplicate rows in the data, which composed over the half of the data. This is the snapshot of the data where user’s id is '5ff1e194b6a9d73a3a9f1052'.

Table

Description automatically generated

1. There are lots of null values in the scanned rewards item list(rewardsReceiptItemList). We have 6941 records of the items in our receipts data. The picture shows the number of the null values of each column. As you could see, most of the columns contain nearly all null values. If we include all this information in our future database, this may possibly lead to inefficiency in the database.

Text

Description automatically generated with medium confidence

Do you know what might cause these issues? In addition, to ensure database efficiency, I would like to know more about the business process regarding creating app users and scanning receipts, which could help me to understand how we collect the data and the meaning of each field of the data. I would also be grateful If you could provide any process documentation and details.

Thank you.

Warm regards,

Yi-Chen