Dear product manager,

Thank you for the information. I have designed a relational database and set up the ETL pipeline using the receipts, users, and brands datasets provided along with two more derived datasets - itemslist and products, for more efficient data retrieving.

I have also gained some amazing insights from our data available, and you could assess those from the SQL queries included in this email.

However, there are also data quality issues discovered:

* Barcode – Barcode has low matching rate among receipts, itemslist, and brands datasets
  + Barcodes are not uniformed in itemslist and brands tables
  + There are 569 unique barcodes in itemslist table
  + More than half of the barcodes (58.14%) are null values in itemslist
  + There are only 89 matching barcodes between brands table and itemslist table, that means that there are only 15.64% of barcode in itemslist table can be found in brands table
* Incomplete users’ information
  + There are 117 users by receipts scanned, but not found in the users table
* One to many relationships between brand ID and barcode?
  + There are 1160 unique barcodes, but 1167 unique brand IDs in the brands data
  + There are 7 barcodes has more than one brand ID

These issues, particularly the barcode issues might raise concerns and extreme inaccuracies in further analysis, the impact could potentially worsen as our production scales.

To resolve and mitigate above issues and to optimize our data performance, these actions could be considered:

* Automate to extract itemslist from rewardsReceiptItemList column from receipts table
* Create an additional product table that includes barcode and description to relate our tables and for better management of barcodes
* Better matched and uniformed barcode data collection
* Unify one to one relationship between brand ID and barcode
* Update users’ information whenever they have a receipt scanned

Please don’t hesitate to let me know what you think and what our next steps will be to optimize our database.

Looking forward and happy to provide more insights!

Best,

Yuxuan