M Gmail

Ferris Deinparvar <fdeinparvar@gmail.com>

Fetch Rewards Receipts Database – Summary & Next Steps

Ferris Deinparvar <fdeinparvar@gmail.com>
Draft

Wed, Jun 4, 2025 at 2:15 PM

Hello All,

I've completed the initial build of the Fetch Rewards receipts database based on the data provided. In addition to constructing the core structure, I've developed a set of reports to address your questions and highlight key data quality issues. I've also created an executable that exports all data quality reports into an Excel format for ease of use.

Below is a summary of structural decisions made during the database design process:

1. Separation of Brand and Product Tables
   The Brand and Product data have been split into distinct tables. This improves clarity, reduces overlap, and allows for easier parsing and analysis going forward.

2. Itemized Receipt Table
   A separate itemized receipt table has been added and linked to the main receipt table. This enables detailed tracking of individual items purchased and their associated brands.

3. User ID Handling
   I assumed that new user IDs will be assigned going forward. However, I've retained the legacy user IDs as they are necessary for data cleanup and reconciliation efforts.

During validation, I identified four key areas of concern and provided next steps for the solution:

1. Duplicate Barcodes Across Different Products

   ○ Please provide a definitive list of available products with their unique barcodes.

2. Duplicate Users

   ○ I've written a script to remove fully duplicated users (i.e., identical records), retaining only one entry. For partially mismatched duplicates, I recommend keeping the record with the most recent creation date and removing the rest.

3. Receipts Missing Product or Brand Codes

   ○ Please provide a list of products and brands along with their corresponding unique barcodes and brand codes.

4. Receipts Missing Associated Users

- ○ These require further investigation. For cases where the correct user cannot be identified, I suggest creating a temporary user record until a resolution can be reached.

Scalability Considerations

The current database design is relatively lightweight and should scale effectively. However, if performance issues arise in the future, we can implement indexing by date or partition the receipts and itemized receipts tables by year.

Please let me know if you have any questions

Best regards,
Ferris Deinparvar