

Director, Data Architecture:

I have completed my initial review of the data provided in the JSON files we receive for 'brands', 'receipts', and 'users'.

These files do provide a substantive basis for data analysis to create relational and analytic models in support of our data strategy. However, I have discovered several issues that need to be addressed to make proper use of this information.

I recommend setting up time to discuss the required project tasks directly.

There are formatting and syntax issues in the data. I corrected these issues in the relational model.

- Field names should not begin with a dollar sign (\$). These are used to indicate variables.
- Field names should not begin with underscore (_). While most databases will allow it, leading underscores often hold meaning in script.
- Words including 'user' and 'role' are reserved in many database platforms and scripting languages. They should not be used as field or table names.

These are some of the issues identified in the initial review:

- We have 'receipt' and 'receipt line item' data in the same JSON file with nonstandard data contents. Since the contents vary, we need to map and validate these fields carefully.
- Recent records are often missing barcodes, which we require for item identification.
- Brand codes are missing from many records.
- The data set contains no 'item' information separate from the brand and receipt data. We need an inventory table.
- The 'category' data within the brand and receipt files are not consistent. We need to create a category list and develop a cleansing process.
- The allowance for users to enter barcodes is a good business process but requires monitoring from the data perspective for proper validation.

We need to build our relational model using best practices and business requirements, then create appropriate data cleansing and ingestion processes to populate that model accurately.

I look forward to speaking further.

Best regards,

Greg Anderson