**Fetch Rewards Coding Exercise**

**Brianna Lytle**

# Pt.1 Review Existing Unstructured Data and Diagram a New Structured Relational Data Model

**Data Tables:**

* Brands (df\_brands)
* Items (df\_items) \*\*suggested new table
* Receipts (df\_receipts)
* Users (df\_users)

|  |  |  |
| --- | --- | --- |
| **Table Name** | **Key Identifier** | **Location** |
| Brands | Barcode | barcode in df\_items |
| Items | Barcode | barcode in df\_brands |
| Items | \_id | \_id in df\_receipts |
| Receipts | userId | \_id in df\_users |
| Users | \_id | userId in df\_receipts |

**Considerations:**

* **Creating a new table (df\_items):** df\_items was created based on the values from the rewardsReceiptItemList in df\_receipts including brandCode and barcode. This was the best approach to ensure receipts data and brand data can be connected. It also allowed a more seamless breakout of different items that may have been part of the same receipt.
* **Further Auditing on Brands identifiers:** For this exercise, we assume that either barcode OR brandCode in the Items table can be used as a linking key to the brands table. However, within our analysis, we found that only a small proportion of these identifiers match to eachother. This may require (i) further understanding of historical receipt scanning process’ and how it may have changed the string of numbers and (ii) identifying a source of truth for brandcodes.

A screenshot of a computer

Description automatically generated