Hi Wade,

I analyzed the data you provided and have some findings and thoughts to offer.

Data Quality Issues:
- Missing Data - There is a sizable amount of null values, which restricts analysis. Only 89 out of 7,381 receipt items (1.2%) have barcodes. This effectively severs the link between the brands table and the receipts table.
- Unclean Data - Nested data structures, like dateScanned, make it difficult to analyze trends out of the gate.

MongoDB Structure:
- I noticed that the data is in a MongoDB format. Although MongoDB works well for application development, it presents challenges for reporting and analytics. Its document-based storage is not optimized for reporting and nested fields like rewardsReceiptItemList make querying complex.

Recommendations:
- Ensure important fields like barcode get recorded.
- Switch to a columnar database. BigQuery, Vertica, and Snowflake are good options to consider.
- To cut down on storage costs, implement a retention policy for tables with large fields like rewardsReceiptItemList.

I have a few questions:
- How many concurrent users pull reports?
- Do you get a large volume of requests daily?
- What are the primary objectives of your team at this time?

I enjoyed completing this analysis and look forward to your feedback.

Sincerely,
Wynne