Hi all,

After reviewing the provided data, I wanted to bring to your attention some data quality issues I've noticed and seek additional information to ensure accuracy and reliability.

Observations and Questions:

1. **Receipt Data:**
    - Many receipt items lack associated barcodes, and some receipts do not have associated receipt items.
        i. Are there specific receipt statuses that would justify this data not being captured?
2. **Empty Values:**
    - Several data fields contain empty values.
    - For example, in currency-related elements like *Total Spent* and *Points Earned*, 46% and 39% of the data is missing, respectively.
        i. Are there established business rules for handling incomplete, missing, or duplicate records?
3. **Top Brand Values:**
    - In the *brands* file, approximately 47.5% of the top brand values are neither marked as **true** nor **false**.
        i. Should we assume that only values explicitly marked as **true** are top brands, and anything else, whether marked or not, is **false**?
4. **Data Intake Expectations:**
    - Are we expecting a steady stream of JSON files, or do we anticipate receiving other file types from the source? Can we batch-process some checks, or is real-time integrity critical?

Recommendations for Data Storage:

As our database grows and processes more files, we should consider optimizing our data storage to better support efficient analytical processing. Currently, the database is designed primarily for day-to-day transactions. While it is possible to optimize our PostgreSQL tables to improve analytical query performance, the increasing volume and complexity of data may eventually require a more scalable solution.

I recommend transitioning to a database like MongoDB as our data requirements grow. MongoDB's ability to handle unstructured data and its high scalability make it an excellent choice for workloads that demand flexibility.

I look forward to your feedback on these recommendations and discussing the next steps.

Best regards,
Brandon