Hey team,

I've been analyzing our transaction, user, and product data, and here are some key takeaways:

🔍 Data Quality Issues:

- **Users:** Some birthdates suggest unrealistic ages (e.g., 100+ or under 10). We should exclude 100+ ages from analysis to maintain accuracy.
- **Products:** Duplicate **BARCODEs** were found. I've applied a deduplication logic for future analyses, but we should review how products are assigned barcodes.
- **Transactions:**
  - **Duplicated RECEIPT_IDs** indicate that some transactions may be double-counted, potentially inflating sales figures.
  - **FINAL_QUANTITY and FINAL_SALE** were stored as strings, with "zero" and blank spaces instead of numeric values. This has been cleaned, but we should enforce strict data types in the front end to prevent inconsistencies at the database level.
  - Many **USER_IDs and BARCODEs** in transactions don't match the Users or Products tables, suggesting potential data integrity issues.

📊 Interesting Trends:

- **Fetch experienced rapid user growth between 2018 and 2022, peaking in 2022 (26,807 new users).** Growth slowed in 2023 and 2024, with new user numbers declining to 15,464 in 2023 and 11,631 in 2024.
  - The sharp rise from 2019 to 2022 aligns with increased adoption of online grocery shopping during the pandemic and possibly successful marketing efforts.
- **Majority of users are 25-44 years old, but spending per transaction is highest among elder users (65-74 years old).**

🚀 Action Proposals:

- **Reassess user acquisition strategies**—Continue targeting 25-44-year-olds through social media, but also explore new ways to attract and engage older users (65+), such as offline events and in-market activations.
- **Address data integrity issues**—Investigate why some USER_IDs and BARCODEs don't match and implement validation checks at the point of entry.
- **Prevent transaction duplication**—Review how transactions are recorded to avoid duplicate RECEIPT_IDs inflating sales metrics.