**Subject:** Data Quality Update, Key Insights and Concerns

Greetings!

I hope you're doing well! I wanted to share a quick update on the data quality analysis for the three key datasets: Users, Transactions, and Products. After thoroughly cleaning and processing the data, I’ve identified a few key issues and insights that I think are important to discuss.

* **Data Quality Issues:** The ‘Users,’ ‘Transactions,’ and ‘Products’ datasets had missing attributes, with additional issues such as overly long User IDs, inconsistent date fields, 335 duplicate rows in Transactions, and 215 duplicate entries along with invalid or missing BARCODE values in Products.
* **Key Insights:** Annie's Homegrown Grocery ranks as the top brand in receipts scanned and sales, with 552 scans by users aged 21+ and $2,321.04 in sales from accounts active for six months, showcasing its popularity and strong revenue impact.
* **Key Concern:** The maximum transaction count per user is only 6, indicating low engagement and infrequent purchases. This may reflect a lack of sustained interest or satisfaction, impacting revenue growth. Enhancing user experience or offering incentives could help boost engagement and transaction frequency.
* **Request for action:** To help us move forward and resolve the outstanding issues, here are the key points that require clarification:

1. Standard policies or methods for handling missing data, and guidelines for selecting appropriate placeholders.
2. Is the long format of User IDs an issue for downstream systems, and should we consider standardizing them?

Additionally, we need to verify whether the user\_ID is encrypted or anonymized, which could impact user tracking. Any information or guidance would help resolve these issues and allow us to move forward with accurate analysis.

Looking forward to your input!

Best Regards,

Kinnera Veerapaneni.