Hi Stakeholder,

I'm Zak from the analytics engineering team. We're responsible for transforming our data to enable data-driven decision-making across the organization. I'm currently working on curating our receipts, users, and brand data, and I've identified some data quality concerns that I hoped you could provide insight on.

After a deep dive into the data in our Snowflake data warehouse, I've found the following discrepancies:

Receipts:

1. Some receipts receive points despite having $0 total spent.
2. A number of receipts have a scan date prior to the purchase date.
3. Several receipt line items have null barcodes or no price associated with an item.

Brands:

1. There are items with duplicate barcodes but different categories/names.
2. Example: Barcode 511111004790 is associated with:

-Category: Baking, Name: Alexa

-Category: Condiments & Sauces, Name: Bitten Dressing

Users:

1. We've identified duplicate users with multiple sign-ups.

Questions for the team:

1. Are these patterns expected or do they indicate issues in our data pipeline?
2. Have there been any recent changes to the system that might have affected data integrity?
3. Do we have current documentation on our receipt scanning process?
4. Are there any known gaps in the receipt scanning process that we should be aware of?

Performance and Scaling Concerns:

As our user base grows, we anticipate challenges with:

* Increased data volume impacting query performance
  + Plan: Implement best practices for materializing and partitioning data
* Ambiguity in field definitions as we introduce more data sources
  + Plan: Establish constant communication between business users and engineers

I'd like to suggest setting up a regular cadence to discuss these data quality issues and our plans to address them. Would you be available for a 30-minute meeting next week to go over these findings in more detail?

Please let me know if you have any questions or if there's any additional information you need from my end.

Best,

Zak Anders