Subject: Initial Data Quality Review and Next Steps for Optimization

Hello [Stakeholder's Name],

As part of our efforts to ensure that our new data model is reliable, I've conducted an initial review to identify certain data quality issues. Here are some findings and the next steps I propose:

Key Findings:

- The Users table contains null values in user IDs and creation dates, potentially affecting user tracking and analysis. The Receipts table has duplicate entries, which may cause reporting inaccuracies. Additionally, the Receipt_Items table includes orphan records, indicating mismatches between items and their corresponding receipts.

Information Required:

- **Data Sources:** Please clarify the primary sources of our data. Understanding where and how data is collected will help us identify the root causes of inconsistencies.
- Are there specific rules or constraints that should always be applied to the data? For example, are there known ranges for prices and quantities that can help us validate data accuracy?

Discovering Data Quality Issues:

I utilized SQL queries to perform checks across various tables, looking for duplicates, null values. This process helps in pinpointing areas where our data does not meet expected standards.

Resolving These Issues:

To address these issues effectively, I would need further details on the business context and any previous issues encountered with similar data sets. Additionally, access to more logs or error reports generated during data entry or collection would be beneficial.

Optimization and Scaling Concerns:

As we scale, the volume of data will increase, potentially amplifying existing issues and impacting performance. I plan to implement automated monitoring and validation checks to continuously assess data quality. Optimizing queries and considering partitioning large tables or moving to a more robust data processing framework like Apache Spark may also be necessary.

Could we schedule a meeting to discuss these findings and outline a detailed action plan? Your insights would be valuable in guiding the next phases of our project.

Best regards,
Sonali Bandi
Analytics Engineer
Fetch