

Hi,

As part of our data analysis process, I've reviewed the current data model and performing an initial analysis, I wanted to share some observations and questions regarding the data and processes. This will help us ensure that the system is optimized for business insights and operational efficiency.

Questions About the Data -

1. **Source of the Data:** Could you clarify whether the data originates from a Point-of-Sale system, an operational database, an enterprise data warehouse, or a shared ledger (e.g., blockchain)?
2. **Type of Data Source:** Is this data coming from a data warehouse, a data lake, or in-house systems?
3. **Purpose of Analytics:** Should we focus on enterprise-level insights (e.g., for BI dashboards) or operational-level analytics for day-to-day decision-making?

Data Quality Issues - By running SQL queries and visualizing the data, I identified potential anomalies like missing or inconsistent values, duplicate records and some outliers in the datasets.

To resolve these issues, we would need:

1. Information about ETL pipelines (tools used, transformations applied).
2. Business rules that define expected behavior for data analytics and dashboarding

Optimization Needs - To optimize the data assets and improve reporting capabilities, we would benefit from:

1. Metadata documentation for all tables and fields.
2. Detailed business rules and normalization for each entity.
3. Guidance on effective data modeling methodologies specific to use cases.

Performance and Scaling Concerns - In production, I anticipate challenges such as:

1. **Query Performance:** As the dataset grows, complex queries (e.g., aggregations across receipts and line items) may slow down. We can address this by indexing frequently queried columns and partitioning large tables by date ranges.
2. **Scaling:** With increasing users and receipts, storage and compute resources may become bottlenecks. Implementing caching mechanisms (e.g., materialized views) and evaluating distributed database solutions could mitigate these concerns.

Let me know if you'd like to discuss these points further or schedule a meeting to align on next steps. I look forward to your feedback!

Best regards,
Sushrutha
Analytics Engineer