Hey Business Leader,

I hope this message finds you well. As I deep dived in the data for better understanding and explore opportunities to improve it, I wanted to share some insights and ask questions

regarding the data to answer the business questions better.

**Questions About the Data:**

To steer the efforts in the right direction, I have a few questions:

1. Are there specific metrics or KPIs that should be prioritised in data analysis efforts?
2. How critical is data quality and accuracy for upcoming business objectives?

**Data Quality Concerns:**

* **Missing Values:** Observed significant missing values in several columns, exceeding 80% in some cases.
* **Duplicates:** Duplicate entries were identified within the data, potentially impacting accuracy.
* **Data Mismatch:** Inconsistencies were found between tables, making it difficult to establish relationships and draw insights.
* **Data Authenticity:** The overall dataset size suggests potential incompleteness, raising concerns about its comprehensiveness and reliability.

This can lead to data inaccuracies, incorrect model building and ultimately, hinder informed decision-making.

**Proposed Solution:**

To resolve the issue, we have come up with a few solutions-

* **Data QC and validation at source:** Apply quality checks and data validation at data source to avoid data quality cascade issues further down the line.
* **Data Quality Awareness for Partners :** Provide data awareness and data ingestion suggestions to partners to improve the data ingestion methods and enhance quality
* **SME Collaboration :** Working with SME would help to better understand the data and help learn their expectations to address needs effectively.

**Information for Optimization :**

While building the data assets for business, optimization is the key for building fast and shipping fast. To optimize our data assets for efficient analysis and reporting, the following information would be valuable:

* **Data Definitions and Documentation:** Clear documentation ensures everyone involved has a consistent understanding of the data.
* **Reporting Frequency:** Knowing the desired frequency of reports helps tailor data pipelines
* **Upcoming Business Objectives:** Understanding your vision for product scaling and future business objectives allows us to align companies data strategy effectively.

**Performance and Scaling:**

As product scales, features will be added and the data will grow exponentially. Keeping in mind, following steps are being taken -

* Transformation Pipeline using dbt to monitor data flow and detect data breaks
* Optimization of SQL queries to improve processing efficiency
* Implementation of separate pipelines for different features to avoid data mixing

Your insights and thoughts about product scaling and vision would give a better idea as we strive to enhance our data quality and optimise data assets for better decision-making. I look forward to discussing these points in detail at your availability.

Best regards,

Sumeet Badgujar