Subject: Review and Analysis of Recent Data Insights

Hello Product Leader,

I hope this message finds you well. I wanted to share some insights and observations from recent data analysis efforts that may impact our decision-making processes moving forward.

In reviewing the data, several questions and considerations have arisen:

**Data Quality Issues:**

During the analysis, we noticed several data quality issues that could affect the accuracy and reliability of our insights. These issues include outdated data, missing values in critical fields such as barcode, final price and potential irrelevant brand codes that do not exist in the receipts.

Please find screenshots below:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Discovery of Data Quality Issues:**

We discovered these data quality issues through thorough examination and querying of the dataset using SQL queries. By assessing the structure, content, and patterns within the data, we were able to identify inconsistencies, and areas requiring further attention.

**Resolution of Data Quality Issues:**

To address the data quality issues, we need to implement data cleaning and preprocessing procedures. This involves standardizing item descriptions, handling missing values in description appropriately, and removing or resolving duplicate entries. Investigate why some brand codes from the receipts do not have corresponding matches in the brand table. Are there discrepancies in the data sources, data transformations, or data loading processes? Additionally, establishing data validation checks and protocols for future data collection and entry can help maintain data quality standards.

A screenshot of a computer

Description automatically generated

Additional Information Required: To optimize the data assets effectively, we would benefit from additional context and insights regarding the data collection processes, data sources, and any transformations applied to the dataset. Understanding the origin and lifecycle of the data can inform our approach to data cleaning, validation, and analysis.

Performance and Scaling Concerns: As we move towards production, we anticipate performance and scaling concerns, particularly regarding data processing and analysis. Implementing efficient data storage, indexing strategies, and scalable processing pipelines will be essential to ensure timely and reliable insights without compromising performance.

In summary, addressing the identified data quality issues and optimizing our data assets are critical steps towards leveraging data-driven insights effectively for informed decision-making. Your input and guidance on prioritizing these efforts and allocating resources accordingly would be greatly appreciated.

Please let me know if you have any questions or if there are specific areas you would like to delve deeper into. I look forward to discussing these insights further and collaborating on our data strategy moving forward.

Thank you.

Best regards,

Ramya