

Hello,

My name is Vinay and I will process your request to provide analytics solutions for your rewards data. When I start looking at the data, I would like to ask you for some clarification to better understand the data and deliver a feasible solution.

Questions regarding Data Description:

- What is the difference between FinalPrice and UserFlaggedPrice in RewardsReceiptItemList?
- Could you provide more information about CPG field in Brands dataset

Questions regarding Data Quality:

While performing Exploratory Data Analysis, I came across following data quality issue:

- TopBrand field in Brands dataset has substantial null entries which becomes difficult to precisely analyze the topBrand information. How do you suggest handling the null values as TopBrand is a significant field and impacts the dataset effectively.
- Missing barcodes in RewardReceiptListItem would make it difficult to retrieve the brand information. Would you like to make updates to capture the missing data?
- More than half of the user's are duplicated. I would recommend adding unique constraints for the users table in the database.
- Many of the fields in the receipts dataset contain null values, and this might have a negative impact when analyzing the transactional data/metrics. Addressing these missing values would help to analyze the data better.

Questions regarding Optimizing the data:

- RewardsReceiptItemList for each receipt is stored in receipt data itself. This requires additional data transformation and processing time. Is it possible to get this data in a separate file? This would save a lot of data processing time.

Questions regarding Performance and Scaling:

- What are your thoughts on retaining the data? Do you need copies of both raw and transformed data?
- What is the frequency of the report refresh?

Please provide me with the following information so that I can decide on the data model, integrity, and structure. Kindly let me know if you have any questions or concerns.

Thank you!