Hello John,

I hope you are having a great day! I recently joined the data science team and have been tasked with designing a relational database for a collection of unstructured datasets. I hope it is alright that I pick your brain with a couple of questions and issues I identified in the data.

1. As a first step, would you help me understand the different touchpoints within Fetch ecosystem where customer purchase data is captured?

   What are the typical issues we encounter when integrating a new touchpoint or data source?

   I identified duplicate records, spelling and case errors, and missing values in the data I am given. This information would serve as a precursor before I reach out to the data engineering team to discuss potential discrepancies arising from data integration.

2. Our business is a three-sided marketplace with customers, brands, and retailers where customers can buy products from brands that are sold at more than one retailer. I wanted some clarity on the concept of campaigns.

   Are campaigns always run by us or do our partners (both brands and retailers) have the power to run campaigns on our platform? This would help me make sense of some of the campaign data extracted by Metabrite in receipts.

3. Every CPG item rolls up to an item category and is owned by a brand. Could you confirm that Barcode, Category Code and Brand Code are unique identifiers for items, categories, and brands respectively?

   I identified brands without brand codes and items without barcodes in the data. It is posing a big challenge in my database design so far.

Lastly, I wanted to highlight that attributes of on receipt such as Item Description and Campaign descriptions are currently tightly coupled with the receipt and lead to data redundancy. This would not scale efficiently. I suggest that since item descriptions and campaign information is constant over extended time periods, they should be moved to their individual schemas in the database design going forward. As a next step, I will work with the data engineering team to achieve this.

I look forward to hearing your thoughts on this.

Best,

Abhijeet Talaulikar