Hi [Product/Business Leader],

I was able to review the sample sales data we've received so far. There is enough foundational info to answer the initial business questions the client had for us, but longer term, there are systemic issues with the data we'll need to resolve internally before we can deliver consistently repeatable data solutions for the client.

- Data quality issues
    - The brands dataset contains a significant amount of placeholder brand names with null values associated. I've filtered them out for now but will work with the upstream Data Engineering team to identify the root cause.
    - More than half the items are missing barcodes. This will impact scalability of the data warehouse since we'll need this info for dimensional modeling.
- Questions on the data
    - The brands dataset includes a combination of brands (Pepperidge Farm, Kellogg's, Frito-Lay) and product lines (Goldfish, Cocoa Puffs, Tostitos). I'd like to align with you on how we define brands and what level of granularity is needed here.
    - It's unclear how brand barcodes correspond to item barcodes from the data, and further alignment is needed. Without a reliable key linking items and brands, it will impact dimensional modeling decisions and scalability of the data.
    - Item transaction data is semi-structured with over 30 unique fields. Further alignment is needed on the necessity of using this data for analytics and what kind of database it should be stored on.

Please let me know if you have additional questions.

Thanks,

Robert