

Hi Jackie,

I hope you all are doing well. I recently joined the Fetch Rewards team as a Data Analyst. I got the chance to quickly perform a couple of analyses on the data to understand how records of users, transactions, receipts, items are maintained in the company and how people are utilizing it. However, working on it I found out a couple of issues in them. I was wondering if anyone from your team could help me understand it in more detail.

1. A lot of data is missing from the database for many fields like brandCode, categoryCode to name a few. If you can provide me with what could be the possible reasons for these kinds of missing data to overcome the possible association of items with the brands and the categories, they belong to?
2. I further noticed that a complete list of items is missing and there is no way we can keep track of the items which we have in the inventory. As a result of which I had data quality issues when dealing with items purchased by customers. In this case, if anyone from your team can help me understand how I can generate the list of items and details related to items.
3. I would like to know how the final price of a total quantity purchased is calculated. As when I was going through the data, multiplying the quantity purchased with the item price it did not seem to match the final price in the data. This seems to be a data inconsistency issue and needs to be looked upon.
4. Lastly, I found that few records are repeated with all details being the same, what could be the possible reasons for this and what can be done to make sure that only one record is being fetched for a particular entry. As each record consists of unique identity to identify them, as records are being repeated, in that case, the uniqueness of records is not verified.

After receiving inputs from your end, I would be able to work with data engineering team on these issues which I am currently facing and getting to know your thoughts on generating data for few tables which are missing and provide an optimal result from the datasets. The only performing or scaling issue I am concerned about here is holding duplicate records leads to more usage of space and reducing the space for other data storage.

Let me know your thoughts on these and I would be happy to schedule a meeting to discuss and elaborate on these points in detail.

Best,

Yuthika Shekhar