

Hi (Person's Name),

My name is Chris Stewart and I work on the Analytics Engineering team. I've been working on a project with data related to our users, brands, and receipts to assist with our analytics efforts and had a few things to run by you.

General Data Questions:

- For the rewards data, there are a number of different price fields. Can you confirm which of these feed into the calculation for the totalSpent field in the receipts data?
- We have data definitions for the users, brands, and receipts data but not the rewards data. Could you send along definitions for the fields in this data set?
- In the brands data, I noticed the cpg fields were nested within a single field but there was never more than one instance of them. Would there ever be a case where a brand would have multiple counts of nested cpg fields and if so should we break this data into its own table?

Data Quality: We processed the raw data using Python scripts and found a few issues after running some basic data quality checks and ran into a few errors.

- Duplicate Records: There are a high number of duplicate records in the user data. Do you know what could be causing this and if there is any reason to keep the duplicates?
- UserIds in receipts but not users data: The receipts data seems to reference a good number of userIds that are not in our user data set. Is there any mapping that needs to be done here of the Ids or is some user data not getting correctly generated?
- Missing Data: There are a good number of fields especially in the receipts and rewards data that are missing values. Can you confirm which of these fields absolutely need values and which are okay having NULLs?
- Bad Role Values: In the user data, the role field should be consistently set to 'consumer' but there are multiple instances of records with 'fetch-staff.' Do you know what could be causing this and if we should delete these records?

Additional Questions:

- Given the high number of fields in the rewards data and how many of them have high percentages of NULL info, are all of these fields necessary or can we save some space on data storage by not loading all of these into our DB?

I appreciate your help with all of this. If you'd rather discuss over a call I can send over a calendar invite for a quick meeting.

Thanks again and hope you have a good rest of your day.

Best,
Chris