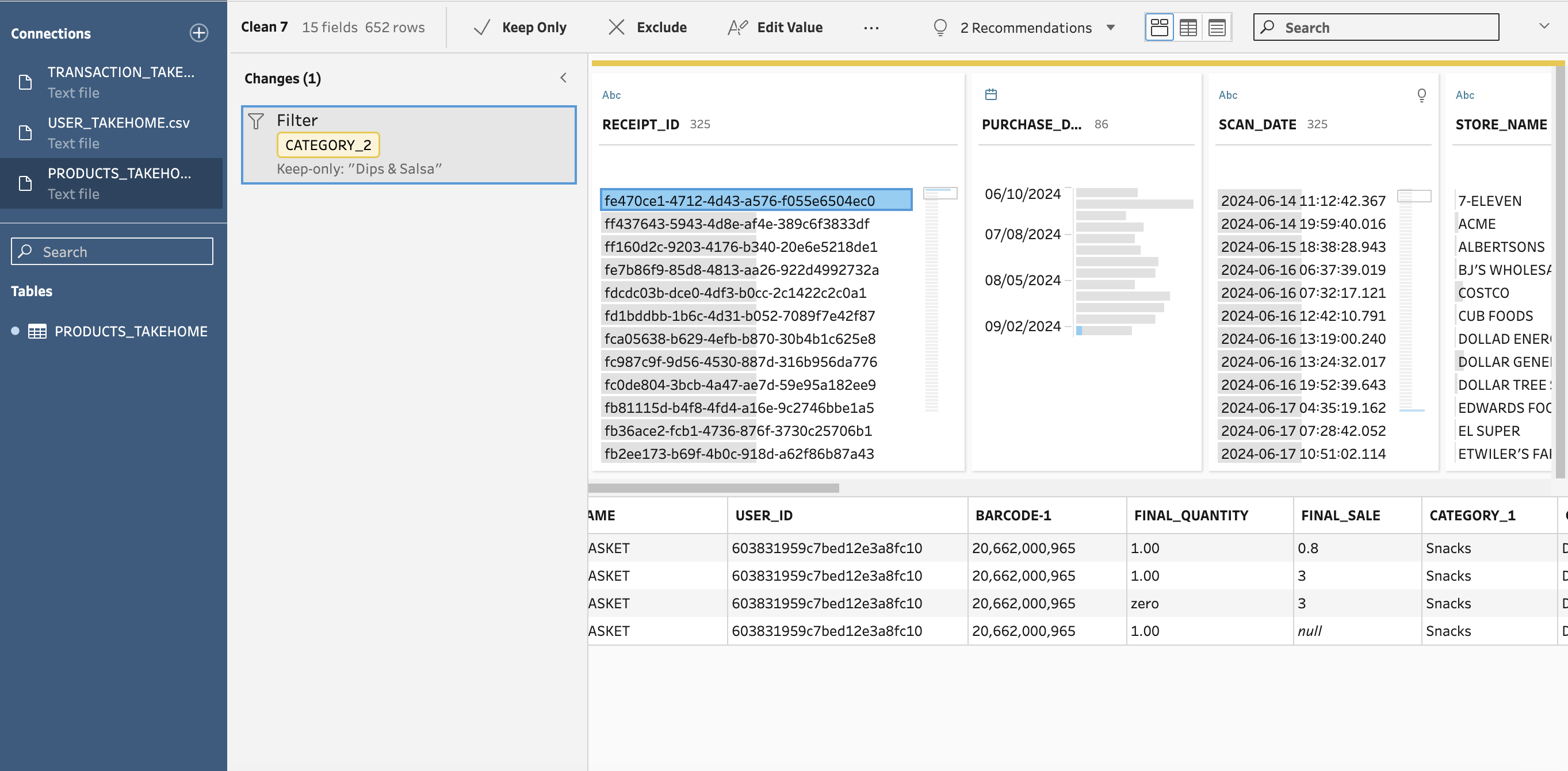
Hello Team,

I hope this email finds you well. I have been deep diving in the dataset recently and would like to share few insights.

First, I found some areas where a bit of data cleanup is needed – like removing duplicates, address redundant values missing values, etc. For instance, in PRODUCTS table, Barcode should be assigned a primary key for each product, but since missing few values it is hard to locate a join record for some in the TRANSACTIONS table for this reason. Filling in these gaps means we get a more accurate view of the customers shopping pattern.

Second, I noticed the USER data is hardly enough, the TRANSACTIONS table has a user\_id field which has more users than the actual user table. I had to exclude using the USER table in one of the SQL queries for this reason, when joining the user with transactions table filters out majority of the transactions data.

Third, the main challenge I had with the data is in the TRANSACTIONS table , the fields- Final Quantity and Final Sale



This scenario here, I noticed these for almost all the receipt\_id ,for the same product , there are multiple rows with different values for these fields. My initial thoughts were Final quantity >=1 means purchased at full price and Final quantity =zero then purchased at discount.

I am eager to work on refining this data and learn and grow with Fetch.

Shreyaa Sridhar