

Hi Stakeholder,

I have spent some time going through and analyzing the users, transactions and product data sets. There are a few things that I would like to point out or get some clarification regarding.

Data Quality Concerns:

- In the User Table, there are many missing record values in the Birth_Date, State, Language and Gender fields. Roughly a few thousand columns for each of these columns except Language which is missing over 30k records.
- The Transaction Table has over 5k missing records for Barcodes field.
- Products table has missing records for Barcodes as well, Manufacturer and Brand columns are missing too for quite some records in this table.

Questions:

1. What is the business use for the Language column in the User table and what would you like to do if there are so many missing records in this field?
2. In the Language Field what does ES-419 stand for?
3. Is the Final Quantity field in the Transaction table the number of products purchased? Why is it often 1 or zero values?
4. Is the Final Sale in the Transaction Table the final price of a purchase? Why is it empty when there is a Final Quantity of 1 and populated when Final Quantity is zero?
5. What is a Placeholder Manufacturer? Is it a real manufacturer or just a fill in value?

Interesting Finding:

Comparing 2023 to the previous year, there was a significant drop by 43% in the number of users who were created. All of the previous years leading up to 2023 had shown significant growth in the number of users who were created.

Kindly take a look through my notes provided and let me know if you could help answer some of my questions. I'd appreciate any guidance or advice you'd be able to provide in order to solve the data quality concerns and gain clarity on some of the business terms I have questions regarding.

Thanks,
Abhishek