Hello Stakeholder!

Nico here from the Analytics Team. I have been running some analysis on the receipts data model and I found some discrepancies that need to be addressed. The image attached is a summary of the dataframe containing information on the receipt items. The first column is the column fields within the dataset, second column is how many entries are not null and the third is the data types. There are several fields here that contain null values which I would like some clarification on to aid my analysis.

For example, I was looking into brand popularity but brandCode is clearly missing a huge chunk of the data with over 4000+ nulls. Is this a normal occurrence? If so, how should I address these null values? I was hoping to link the receipt items to the brand table get more information but I do not see a similar key to connect the two tables.

I would very much like to know more about the business process and how it ties to the data model. This would allow me to better understand the fields and how they are flowing into the system. It would give me a better idea on how to handle them in the future. Would appreciate any insight you can provide!

Here are some questions I prepared ahead of time so we can have a more productive conversation:

1. What is the current business process? How does it relate to the different fields in the model?
2. What is the connection (common key) between receipt items table and brand table?
3. Which fields should have null vs no null?
4. How do we address the fields that are null? Is it something we can handle with data manipulation or is it better to fix it at the source?

Table

Description automatically generated