Hello Stakeholders!

I hope you are having a quick Friday and excited for the weekend.

Below you will see my insights into the data that has been given.

* The data quality has several different areas of concern. The user information has blanks, where I believe the data should have a default value of some sort if not entered just for the ease of filtering out. The dates should also be consistent format as well as not have a Z at the end. With the Products, there should not be an empty category1. All items should be able to fit into some category or again have a default value. The barcodes are also a point where they should all be the same digit amount. At the highest being 14 then the rest should have leading 0’s. With the transactions, the zero entry in the final quantity but no final sale price matches up to the no final sale quantity but sale price. It is as if there are two entries for each receipt but they aren’t one entry because of this issue.
* The most interesting thing I found in the data except that you have no users who were born in 2019 is that when it comes to barcodes that have been scanned, Coca Cola is leading the way by almost 300 but the staggering thing is that over 700 receipts come back to a null brand giving no insight into what people are purchasing.
* When it comes to help for these issues, making sure there are proper and documented processes in place helps guarantee a better data solution. Whether its consistency or better data, knowing how the data interacts with each other is important. Correctly uploading the data can be one of the biggest reasons for data integrity issues (Which is what probably happened to me).

Thank you for your time and I hope you have a great weekend.

-Royce Salay