Fetch - Data Analyst Take Home - Third Exercise

Dear Stakeholders,

I have reviewed and analyzed the three files provided in an effort to provide some insight into your sales. However, I came across several data quality issues that hinder my ability to tell a complete story.

1. Of the 50,000 transaction records provided, all of the Receipt IDs are duplicated. Only 24,440 unique Receipt IDs exist. Does each row represent a separate line item on a receipt, or is there a chance we have some of the same transactions accidentally repeated multiple times?
2. What does the Final_Quantity column represent? I can't tell if the quantity represents the quantity sold, the quantity remaining in inventory, or something else. I see several rows where the Final Quantity is zero but they have a sales price and vice versa.
3. What does the Final_Sales column represent? I can't tell if this is the total price, the unit price per item, or something else.
4. It appears that the Product List may also be incomplete. There are 4,465 barcodes on the Transaction List that are not included on the Product List. In addition, some items on the Product List are missing key information like Category and Brand.
5. I found 4,305 transaction rows missing barcodes on the Transaction List, so I cannot identify what the customer was purchasing
6. It appears that the list of 100,000 users may be incomplete. I identified 17,694 User IDs on the Transaction List. However, only 91 of those User IDs could be found on the User List. This means that the close-ended questions from exercise 2 only include the results from those 91 users. In addition, one of those users is missing a birth date which decreased the users to only 90 for the first close-ended question.

I would greatly appreciate if you could provide some direction on how to resolve these, so that I can provide you with accurate information. In the meantime, I've constructed some visuals to give you an idea of what kinds of insights we can provide with this data. I found that new users have grown tremendously since inception in 2024 to a peak of 26,807 new users in 2022. However, growth dropped off significantly in 2023 and 2024.

Please let me know how to proceed. Also, please let me know if you have any questions, comments, concerns, or other feedback.

Happy Wednesday!

Lisa Armetta