

Hi XYZ ,

Good morning. Hope you are doing good.

I am writing to discuss my current Data Analysis project. This project is about analyzing data and exploring more about receipts, brands, and users.

While working on this project, I came across few data quality issues i.e., duplicated data records, missing data etc. I worked on data cleaning and preprocessing techniques to remove the redundant data. However, I have few questions regarding the data.

1. Why do the receipt product [barcode number 4011] has a description as “Item not found”?
2. Can you explain more about Total price calculations? Do we add prices of the product with barcode 4011 directly to the total price?
3. Why do we see receipts for total spent amounts when there are 0 total purchase items?

Request you to help me regarding the questions above.

Also, I would like to check a data directory file which contains description and the data attributes details.

Please find attached files/programming notebooks for your reference.

Thanks for your help and time.

Best,

Prathamesh Limaye