Hi,

I hope you are well. I'm writing regarding data quality issues on the Brands, Users and Receipt data provided for user behavior analysis.

When I processed the unstructured data and generated queries to answer some business questions, I discovered some data quality issues as follows:
1. After dropping duplicates, there are only 212 users in the Users table. There are 117 records of distinct users in the Receipts table but these users have no information in the Users table. The Users data should be collected and completed in the future.
2. The Receipts data of March in 2021 is not enough for analysis so the data should be updated in time.
3. There are lots of null values in the data tables, resulting incomplete information problem for analysis.

To resolve the data quality issues, I need more information about the data collection. I'd like to know:
1. where this data is from and how it is generated
2. how often is the data collected and updated
3. is it possible that missing values appear because of encryption or any exporting problems

Additionally, I have some plans to optimize the data assets. I'm wondering if you could provide me more details about the data features. To understand which features are more important would be useful for business analysis.

Finally, with the increasing of data, our team need to optimize the ETL pipeline and make sure that the data can be collected and saved completely and timely.

If you have any questions or concerns, please don't hesitate to let me know.

Regards,
Shirley