

Subject: Questions and observations regarding data provided

Dear Stakeholders,

I reviewed the data sent to our team and have a few questions regarding the data that I would like to ask before loading or designing our data warehouse. I discovered these data discrepancies by evaluating the null values in the data and by observing the data closely.

The Users data has duplicate rows which is creating noise in the data. The role field seems to be redundant as its always set to 'consumer'. Are there going to be future additions to this field in the future? The receipts data seems to have dates in a UNIX_TIMESTAMP format, is this expected to be in this format? Some of the data types are not in appropriate formats. Are the data types mentioned expected to be in this format?

We would be using brandCode to identify and link brands to receipts data. This field has missing values which would make it difficult to link the scanned item with the brand name. Is it possible to populate this field to its best extent so that many of the business queries could have better results. pointsEarned has missing values which makes it difficult to maintain the points earned through receipt for a consumer which is an important aspect. This needs to be investigated deeply to understand the impact. lastLogin, state and signUpSource have null values. Since lastLogin date is missing for many consumers, it would be difficult to identify if these consumers are still interacting with the app and when the points would expire. state has null values which would make it difficult to understand the consumer's demographic information for marketing and analysis by location.

I did not come across any major scalability issues regarding this data, but the efficiency can be increased while creating the database by using optimal querying strategies.

Let me know if you have any questions or concerns regarding the above information. Hope to hear from you soon.

Thanks,

Niharika Apte

Data Analyst