Hi X,

I hope this email find you well. As I am working on the three data files of the company that is brands, receipts, and user, I have few concerns and questions regarding the data. Below are few questions I have about the data:

* 1. What is the data source of the data?
  2. Why is it particularly in json format?
  3. What are we looking into the data (at a granular level)

While processing the files, I even came across some data quality issues like null values in certain columns, barcodes of two different items being the same for which I have also attached snippets. I came across this issue while I was on initial phases of data transformation and checking the data in SQL Server Management Studio. Another issue is of duplicate records that existed in very important \_id columns for brands and users. For the receipts scanned, there were many null values for id column of user. That is identified as potential data quality issue as there must be a user information associated with all the receipts and transactions that are happening.

Graphical user interface, text, application

Description automatically generated

Graphical user interface

Description automatically generated

To solve this issue we can start to get a better idea of the data source first and then what other data we can use to fill out the missing values in the columns to decrease the discrepancies in data-model and database reports. Another way is to drop out unnecessary columns which have a negligible impact on the database reports.

In my opinion, we do need some more information to further optimize the data assets and maintain a normalized database schema. To achieve this, the team can take time to understand the data better, build a data-dictionary for new columns and work on maintaining uniqueness of each record. However, there would still be some concerns around the performance and scaling and one of them can be discrepancies in reports due to filling out null values with some default value or zero and another if the data sources being disparate or not.

Please do let me know if you have some thoughts over the data and would like to discuss further about it. Please feel free to reach out if there are other reports you would want to see regarding brands performances and active users.

Regards,

Nishtha Bhakta