Hi [Leader Name],

I just finished working on the Receipt-User-Brand data set, and I have some progress to share with you. After reviewing the raw data, I would like to propose this new structured relational data model [readme figure 1]. In the jupyternotebook [Fetch_Reward.ipynb], I parse the json files into the following five tables: *receipts, rewardsReceiptItemList, users, brands*, and *cpg*.

I wrote a SQL query to compare *average spend* and *the total number of items purchased* from 'rewardsReceiptStatus' that were 'Accepted' or 'Rejected,' respectively. However, the 'Accepted' label is not present in the data set. So, is it acceptable to use 'Finished' and 'Flagged' as accepted labels instead? I checked these are the only two labels containing both *spend* and *the number of items purchased*.

I noticed that rejected *spend* and *items* are relatively lower than those were accepted but still significant. So I decided to relate the receipt table to the user table to see if there was a user reaction due to rejection. In the 'lastLogin' column, 20 percent of the data was missing. There could be a data entry error, but I suspect it could be due to the user not being active when the data was collected. I noticed there was an active user column that FetchReward flagged. I wish to know the rules to determine if a user is active or inactive. If missing values are related to possible user inactivity, I will need to know the March 2021 data to determine whether these null values are the monthly user churn of February 2021.

It would be helpful to connect with someone from the software development team to confirm the reason for the missing data.

I am concerned with the user retention rate and would love to help improve it.

Best,
Jason