

Hello fictitious stakeholder!

Thank you for the opportunity to work with this data, and I hope you are having a good week so far.

I have been able to make steps toward determining answers for your stated business questions. I was able to determine the average spending for customers whose receipt was either rejected or accepted by the rewards system, as well as the total number of items purchased in each category. The table below provides that information:

	average_spend	total_items_purchased
ACCEPTED	80.854305	8184.0
REJECTED	23.326056	173.0

As for the other business questions, there are items I would like to chat with you about that will help me in perfecting your data as an asset and enable these questions to be answered confidently, plus open the door to other avenues of analysis. Let me know when you are available this week!

To give you some heads up, most of the questions revolve around how the data is sourced and what is desirable to keep stored for later access. If there is uncertainty in what to retain, I may be able to make suggestions after a discussion on what the data is primarily used for, or what the primary business objective is.

To be more specific, I'll list a few questions that often contain elements from both categories.

1. To refer back to some of the questions your business has asked, many of them require information on the brand of the sold items on the receipts. With some investigation, I was able to find, for some of the receipts, information on the brand of the item, as well as its barcode. However, out of the 2600 items with brand information for 227 unique brands, only 42 of these brands are currently in the brands database, and out of 3090 item barcodes, none of the barcodes match brand barcodes in the brands database (though they might match item barcodes in a products database; see below). So, using the database to get information on brands would not be representative of all the items sold. The information on the brand only occurs within a tough to access structure and only sometimes, besides less than visible, so if this information is desirable changing the way this data is collected and stored will be useful.
2. The column descriptions for the provided data references other tables, such as partnered products and consumer packaged goods (CPG). Is it possible that I get access to this information?
3. The data is stored in a format that I am unfamiliar with and would need a method of translating to normal date format. Do you know where I can get that information?
4. One of the files refers to a bonus point schedule. Are there multiple schedules, and what are the stages? Learning more about the way bonus points are earned and how many bonus points users have seems interesting, as there may be a way to correlate how the different schedules and methods of earning points affect both customer retention and impact your business.

As you can see, these questions and more are based in improving the quality of the data in the database such that it makes it easier and more efficient to retrieve, and can easily be understood by new analysts accessing the data. I would like to make sure the essentials are recorded while leaving room for growth by maintaining information that could be useful in the near future. As the amount of information scales, it becomes harder to deal with issues in data integrity. So, please provide some availability so to help me better understand your requirements so I can get this right for you know. If you feel there is someone more familiar with specific parts of the data process, I would be happy to be connected with them as well.

Feel free to ask questions regarding the table I shared too! Relatedly, I would love to know how best to share information going forward.

Thank you for your time,

Matt LeGro