Hello *Business Leader*,

I have just finished passing through your data set and have done my due diligence to answer a few of your questions, barring a couple of questions of my own. But first, let's discuss the good news. Regarding the following two questions, I have come to the following conclusions:

- **When considering *average spend* from receipts with 'rewardsReceiptStatus' of 'Accepted' or 'Rejected', which is greater?**

  I had to make an assumption regarding the labeling of your dataset, "Accepted" is not a status found in the dataset you provided me. Having noticed this, I noted "FLAGGED" was the status most likely to represent "Accepted". Can you confirm whether or not this is the case? If it is not the case, I will need to rerun my queries to verify the information I am about to give you is correct. Having said that, the average spend for a user that received an "Accepted" status was $180.45, and the average "Rejected" spend was $23.32, the difference being $157.12, just incase you needed the quick math there.

- **When considering *total number of items purchased* from receipts with 'rewardsReceiptStatus' of 'Accepted' or 'Rejected', which is greater?**

  My conclusions here follow my reasoning from above regarding "FLAGGED" and "Accepted". "FLAGGED" returned less items purchased in total than purchases that received a "Rejected" status. Without the full considerations for your rewards program, I am interpreting this result as indicative of a condition that require a user to spend above a certain amount on a single item in order to receive the "Accepted" status for the rewards points to be granted.

On to my own questions, the biggest issue I ran into when trying to solve for your other questions was the representation of the dates in your datasets. Without being overly technical, your dates were represented as serial numbers, rather than in YYYY-MM-DD or a similarly recognizable format, and I was unable to parse these representations back into a usable datetime format. Do you know, or would a member of your team know, if your dates use zero-padding? The answer to this question could prove helpful in answering your other questions.

The last concern I ran into was the uniformity in serialized IDs across your three datasets. The id columns in each dataset are following a very similar structure, and this could lead to a scalability issue as we move toward production and start to consider adding more relationships across the datasets. As an example, it might be helpful if we created a relationship between the "Brands" dataset and the "Receipts" dataset to more easily answer your first two questions of:

- **What are the top 5 brands by receipts scanned for most recent month?**
- **How does the ranking of the top 5 brands by receipts scanned for the recent month compare to the ranking for the previous month?**

  Just food for thought.

For now, I'll await your confirmation on my questions regarding whether or not "FLAGGED" can be interpreted as "Accepted", and whether or not your datetimes make use of zero-padding. Once we have answers to those questions I will rerun my original queries if necessary, and I'll get to work answering your other four questions from these three datasets. Once completed, I'll compile all of the findings into a much more concise email for your recording keeping and referential purposes.

Thanks for working with me,

Hamilton