Email Background

For clarity, I am using the fictional name Andrew as being a product manager. Along with this, Charles will be the fictional data engineer. For this email, I am operating under a fictional event where I am working on a data request from Andrew. This request outlines a need to have a data table built that will continuously track the users, receipts, brands, purchased items count, categories, total spent, total bonus points earned, total bonus points earned reason, total points earned, reward status, and all related date fields. The table is designed to help Andrew better tailor Fetch's promotions and rewards to its users.

Good Afternoon Andrew,

I hope this email finds you well and you are having a great week. I wanted to touch base and provide an update regarding your recent data request.

In my typical process of developing analytical models, data quality is the first step. Unfortunately, I have come across a few quality issues within our source tables that I will need to address before moving forward. From my understanding of your request, below is what you will need along with the quality issues I am seeing:

- Brands: This will help you better understand which brands are most popular. Within our source table, there are many brands that are missing and would provide an incomplete picture.
- Categories: To better understand what group of products our users are buying most. Like the brands issue, there is a lot of missing data within the category column.
- Receipts: To know what each purchase consisted of. As of now, I do not see any issues with this data.
- Purchased items count: To quantify the brands and categories the users are purchasing. This is another issue with incomplete/missing data and therefore wouldn't allow us to quantify the data properly.
- Total spent: Understanding how much our users are spending on average for each receipt along with in total for a set amount of time. This will empower you to create promotions that would best incentivize our users to continue using the platform. Once again, our source table has a lot of missing values.
- Bonus points earned and the reasoning: Understanding how many bonus points (and the reason) our users are receiving can help understand how they are using the platform. From what I can see, there are currently no quality related issues with these fields.
- Total points earned: Points are a key field to understanding our user's behavior as this is the reason people use the app. Accurately keeping track of the points our users are

earning is key to providing a positive experience on the platform. Once again, we are missing a lot of values within this field.

- Reward statuses: Having the ability to see whether a receipt was accepted or rejected. I am currently not seeing any "ACCEPTED" status; however, I see a "FINISHED" status. I will need to talk with Charles to understand if there is an "ACCEPTED" field or if the "FINISHED" field is meant to be it.
- For all the date fields, I am seeing missing values in some key columns. Along with this, I don't believe the date format is ideal and needs to be adjusted.

To provide the most accurate and reliable dataset for you, I will need to first address these quality related issues. I believe these problems can be addressed by reviewing how we are pulling the data from the source system. I have already reached out to Charles and have a meeting scheduled this week to review everything. After our meeting, I will provide an update along with an updated timeline for when this data should be available.

Please let me know if you have any questions or concerns or would like to get together and discuss further.

Best,

Christian