Hi (stakeholder)!

I hope you're doing well! I've just finished preprocessing the three JSON files provided to me and wanted to share my initial thoughts and findings.

During the preprocessing period I noticed a few data quality issues that need attention in order to move forward with leveraging this data to improve business outcomes. I've taken the liberty of breaking these data quality issues down into a prioritized list in hopes of streamlining our call to action.

1. Most notably, the receipts JSON file (which includes the most granular transaction data available to us within the nested rewardsReceiptItemList table) contains excessive missing values. These missing values would be detrimental to any insights drawn from the data in its current state. How are we collecting this data? Can we explore the fields of this dataset together to determine if/how each one still serves our business goals? Understanding where and how this data is being collected will help us determine action items for improving accuracy and efficiency— which ultimately leads to better insights!

2. There are 495 entries for user IDs in the users table yet only 212 are unique values. This violates primary key constraints and would lead to data integrity issues when building/querying a database.

3. The file type (JSON) for the data is not optimal and requires additional processing time that could be utilized in the analysis and story-building phase. Are we open to building a data warehouse in order to streamline analysis? Having a warehouse for this data to live in would cut back on processing time as well as improve data quality and consistency. As the business scales, I believe a data warehouse would become a necessity to ensure peak performance!

I suspect these findings and questions will lead to further conversations regarding the data we are collecting. I'd love to set up a 1:1, at your discretion, to discuss these further so that we may determine the best course of action. I look forward to working together and hope you enjoy the rest of your day!

Best,

Timothy Wardlow (he/him/his)
Data Analyst
Fetch Rewards