To: "stakeholder@fetch.com" <stakeholder@fetch.com>

Fom: [tommccarthy@outlook.com](mailto:tommccarthy@outlook.com)

Date: 03/04/2025

Subject: Follow up on Data information request

Stakeholder,

Thank you for sending these files for review. Your answers are included. There are a few issues I would like to highlight. If you would like, we can set up a date and time to meet and perhaps include the data team to review and resolve them.

Issues:

* The Brand file you sent has a column called "brandCode". This code comes from an unspecified product file. If you need this data point, please send the associated product file. If not, just take this for information only.
* There are a variety of barcodes in the data in differing sizes. Unless otherwise noted, only 12 and 8 characters are valid. Please review them.
* There are a high number of duplicate Receipt ID’s. Is this expected? If they represent multiple Receipts for the same overall transaction, there must be a way of connecting them. Please review.
* There are several other issues I’ve discovered that I’ve included in a word document. Please review them.

I would like to point out that while these are small sets of data, if we push them out at scale, we will see real performance issues. Some suggestions:

1. We need relationships between data that are integer based. Right now, they are alpha numeric, which is very slow.
2. We need to add indexes to the data to improve performance and reduce bottlenecks.
3. At scale, it’s possible we could process 1mm receipts a day. We should consider off loading or archiving older receipts to improve overall performance.
4. We should look at how to proof the data before posting it. There are a lot of inconsistencies in the data that need to be addressed now and prevented in the future.

I found no Brand transactions for the last two months while completing your "Top 5 Brands" requests.

This might be because there are several barcodes on the monthly receipts that do not have a matching barcode on the Brands file. An audit of that relationship might be in order.

The answers to the remaining requests are below:  
  
When considering average spend from receipts with 'rewardsReceiptStatus’ of ‘Accepted’ or ‘Rejected’, which is greater?  
**Answer: Accepted, with a spend amount of $894.96**  
When considering total number of items purchased from receipts with 'rewardsReceiptStatus’ of ‘Accepted’ or ‘Rejected’, which is greater?  
**Answer: Accepted with 295,613 Items purchased**

Which brand has the most spend among users who were created within the past 6 months?

**Answer: Cheeto’s with $4721.95 in spend**

Which brand has the most transactions among users who were created within the past 6 months?

**Answer: Swanson with 3. \*\*\* This result is most likely influenced by the issue above related to barcodes. We may want to re-run this after that audit/fix has been done.**

If you have any other questions or issues, please feel free to reach out.

Kindest regards,

Tom McCarthy