

Hi [Stakeholder],

I hope you are doing well! I am working through your analytics request and I have a few questions along with some updates regarding the data.

Part of the analysis process for requests requires a data quality review and I have a handful of concerns about the data that if used in its current state, would result in a skewed analysis that would affect business decisions, both, internally and externally. For example, there is a significant amount of missing data, duplicated data, and data that looks to be incorrectly assigned.

Summary of data quality issues:

- Duplicated data:
 - There are only 212 users, with 495 user records, meaning the data would be overstating the number of users we have significantly.

	totalUserRecords	uniqueUsers
1	495	212

- Incorrectly assigned data:
 - The brand 'Caleb's Kola' (along with other brands) is seen to be assigned to two different brands under two different cpq's. We will need to confirm which is the correct assignment to ensure the cpq's are getting credited correctly.

	barcode	brandName	brandId	cpq_id	category	categoryCode
1	511111000518	Caleb's Kola	5a4d23dae4b0bcb2c74ea77e	5332f5f6e4b03c9a25efd0ba	Beverages	NULL
2	511111004912	Caleb's Kola	5d601d74a3a018514994f422	53e10d6368abd3c7065097cc	Snacks	NULL

- Missing data:
 - Roughly 50% of the category data is missing. Where this data is not significant to all analyses, future projects will require this to scale our efforts to enter into new categories.

I have put together an in depth overview of the issues that I brought to the analytics, operations, and engineering teams to further understand the underlying issues and how these issues can be resolved.

- If you would like to see the in depth overview, please see:
[fetch-rewards-data-analyst-assessment](#).

If you'd like to have a call to review the above, I'd be more than happy to! I will continue to keep you updated on the timelines to resolve these concerns.

Please let me know if you have any questions.

Thank you,

Brian Miller