Hi Team,

I hope you’re doing great. I just had a few queries regarding the Rewards data I was trying to work on. Please let me know a convenient time we could discuss this. Here are some of the queries I had.

**What questions do you have about the data?**

How are these data points collected and the data source?

Can you please give me overview on the process of how a user gets rewarded?

**How did you discover the data quality issues?**

I have discovered some data quality issues on my preliminary analysis and found that barcode values for brands are not consistent and this is leading a bad data quality.

**What do you need to know to resolve the data quality issues?**

I just needed to know how the barcode data about brands have been collected.

What is meant by User flagged barcode?

On what basis does a receipt gets rejected?

What happens when a receipt gets rejected?

Can you explain me what some of theses data points mean in a business sense?

**What other information would you need to help you optimize the data assets you're trying to create?**

What are business problems were trying to solve using this data?

Is the end goal of the data being to run analytics or use it in an application?

How much data is being generated? How many new users are signing up daily?

What is Fetch Reward’s business model?

**What performance and scaling concerns do you anticipate in production and how do you plan to address them?**

As were dealing with receipt and rewards data, no real-time/streaming data processing is required. The data size will grow as more user’s signup and more offers we run. In that case we may need better processing. Maybe distributed processing with Spark could be a choice.

To store historical data lakes could be used. More cloud-based solutions could be used rather than in-house solutions, which are harder to scale.

Also, I think better data collection could be possible. If we are running analytics or doing BI on this data, we could create a dimensional model on a warehouse like Redshift which could reduce the data query time.

I know this is long, please let me know if we could set up a call which would be so much easier and quicker. 😊

Thanks,

Sri