# Question - What questions do you have about the data.

## **Answer** - I want to make sure that the information provided, is what you’re looking for or close to it, and if there are further questions about the data pulls. I did run a QA across the data and found that a lot of barcodes are missing from our brands data which can cause an issue in getting accuracy of what our user base purchased brand wise.

# Question - How did you discover the data quality issues.

## Answer - I discovered the data quality issues while running through the questions asked, ran a few QA queries against the data to see that we are missing a good chunk of barcode data. Something i’ll need to look at is where/why this data is not being recorded and if it is an issue with the data I had ingested or outside of my control, in-which i’ll need to reach out to another team for help.

# Question - What do you need to know to resolve the data quality issues?

## Answer - I’ll need to know where the data is coming from, and which team is giving us that data, so I can request help if needed to resolve the data quality issues or if I need to reach out to an engineer to fix the data issue if its out of my control.

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

## Answer - Would need to figure out how to automatically ingest the data assets instead of manually ingesting myself. Also taking a look at to see if we truly need some data or if it can be left out to save on costs, performance and optimization. Seeing what is the best visualization data tool, to be able to setup metrics to analyze from to spot any data issues easier.

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

## Answer - I expect performance to be slow when getting the receipt data, as there can be many products purchased per receipt, which can slow down performance. I can create different data sources to ingest the data and separate basic receipt information and receipt product information. This will help performance, and help with scaling down the road.