**Communication with Stakeholders Q&A**

The below information can answer the questions about data quality issues and on how to address them:

1. What questions do you have about the data?

* The main question I had when I started building the relational model diagram is how do we link the Receipts and Brand tables. I anticipated that the field barcode can be used to connect both the tables but when I tried to analyze the data as shown in Data Quality Issues.py. I realized there is no similarity or common data in barcode field between the tables Receipt and Brand.
* This problem was solved when I found common data between both the tables using the field brandCode as shown in the python code.

1. How did you discover the data quality issues?

* As mentioned above, when trying to build the relational model diagram, I found difficulty in linking the tables Receipt and Brand.
* So, using Python Pandas, I debugged and analyzed the data and finally found the relation between both the tables.
* The other data quality issues were most fields like brandcode, purchasedDate were empty. The field brandCode has been used to link the tables and so it should be empty.
* This has also been identified while understanding the data in Data Quality Issues.py python file.

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

* I wanted to know if there is an alternative or a way we could link the tables Receipts and Brand directly. Because most brandCode data is empty but is an important field for linking the tables.

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

* The date format in json files look complicated and have to be converted to an understandable format which I have done using pandas as shown in Data Quality Issues.py.
* A little more information on date format can help understand and query the data better.

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

* The date in the fields like dateScanned, createdDate etc look complicated and are not understandable or in readable format.
* So, they have to be converted to simple format which I have done using pandas as shown in Data Quality Issues.py which can be a problem in production but has been resolved.
* I have converted the json file to csv for easy importing of data and to easily understand or write sql queries.