Hello Stakeholders,

After modeling the data that I received, I had the following questions.

- How often is the data being collected?
- What are some of the goals for this data?
- How many users do we anticipate having?
- Do we plan on opening it in other countries?

When reviewing the Receipt Item List data I found that the majority of the information had 80% - 100% of the data missing from the records. See the attached report for the breakdown. <data-analysis-rewardsReceiptItemsList.csv> I believe we should look into what is causing this information to be missed and determine if it is actually needed. When looking at the users.json file I noticed right away that there are duplicate records for the users, creating a Primary Key on the Users table would help prevent these duplicate records. Additionally, when attempting to create a relationship between the Brands data and the rewardsReceiptItemList I found that using the brandCode was the best way to do this. I ended up finding 234 records where the brandCode was null. Next, Huggies and Goodnites both had 2 records with the same brandCode and one of their records had topBrand set to 'TRUE' and the other set to 'False' making these data points inconsistent. Furthermore, we could remove all the records without a brandCode and determine which of Huggies and Goodnites are accurate in the database.

In order to best optimize these data sets, I would need to know the following.
- What data is collected?
- How the data is collected.
- How the data is being entered.
- Is the data that is entered complete?
- When is the data collected?

Some of my concerns when it comes to scaling and performance are as follows:
- The total size of the dataset.
- The total number of distinct datasets that should be related to one another.
- The frequency at which the data is updated.
- The frequency at which the data is queried.
- The relative difficulty of processing queries and/or updates.

Finally, you can improve your data scalability by applying transformations wisely and allowing yourself the flexibility for future change.