NOTES

After review the json files, I decided to go with the relational data model as seen below. The brands and users dataframes were straightforward in their creation. For the receipts dataframe, I had to make a decision as to the structure of the table. Specifically, the rewardsReceiptItemList field contained a list of dictionaries, with each dictionary containing information about each item purchased on the receipt. To make the table more digestable, I decided to convert the format of the table where each item on the receipt contained its own row. I also had to create a new primary key that combined the receipt id ('id') with the item number on that receipt ('partnerItemId'). To join the tables, the 'id' field in users joins on the 'userId' field in receipts, and the 'id' field in brands corresponds to the 'rewardsProductPartnerId' field in receipts. Code used to explore the data can be found in the json_read.py file include in the repo.