Hi,

I have gone through the brands.json, receipts.json and users.json file and have done some analysis. I have some questions about the data:

Is it possible for each record in receipts to have brandCode? Currently only few records have brandCode

Is the barcode in rewardsReceiptItemList of receipts.json file barcode of the item?

I have found couple of data quality issues. I found data in users and brands files to contain duplicate data and missing values. Through my initial data validation using python I found that 283 records out of 495 records in users file are duplicate records. signUpSource and state fields of users file contain many NULL values. Also, there are missing values of brandCode and categoryCode in brands file. To resolve the data quality issue of missing values I would need extra information regarding users and brands. I also wanted to know does users file also contain information of fetch-staff who are also customers. It would be great if I can get some information about cpg\_id. One of the performance issues I anticipate is keeping track of brands information in receipts, since not all receipts records have brandCode which give brand information. To resolve this brand\_id can be used in receipts data to better identify brands information.

Thank you,

Mohit Venkata Rama Swamy Korai