Hello,

It has come to our attention that there are some data quality issues found in the datasets which pertain to the receipts that have finished processing yet show no items in the itemlist column. One of the probable explanations to this is expected in the OCR upload from users phone to database pipeline. Although these are just a couple of cases, if there are more in the larger pool of data, we might have to do a reupload of all the receipts that have points rewarded to the user but show no items as we may be losing partner information that may be contained in those receipts.

As receipts grow into billions, the team anticipates performance and scaling concerns which can be addressed by writing efficient SQL queries to fetch data, indexing the tables as well as adding nodes to scale as demand for storage grows.