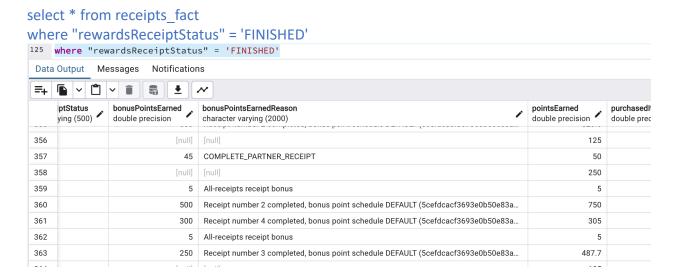
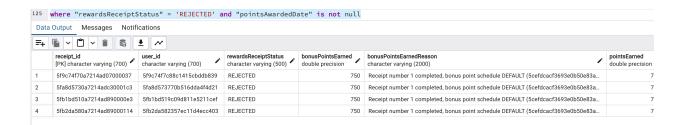
Receipts DQ -

ALTER TABLE receipts_fact
ADD CONSTRAINT pk receipts fact PRIMARY KEY (receipt id);



There are some NULLS in bonusPoints even when status is Finished and pointsawardeddate is also present. Is this behaviour expected?

select * from receipts_fact
where "rewardsReceiptStatus" = 'REJECTED' and "pointsAwardedDate" is not null



Is the above behavior expected? 4 records where the status is rejected still points awarded date is populated.



ALTER TABLE receipts_line_item

ADD CONSTRAINT pk_integration_id PRIMARY KEY (integration_id),

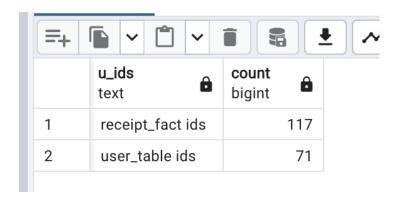
ADD CONSTRAINT fk_receipt_id FOREIGN KEY (receipt_id) REFERENCES receipts fact(receipt_id);

select * from receipts_fact
where receipt_id not in (select receipt_id from receipts_line_item) and "rewardsReceiptStatus"
= 'FINISHED'



We observe 2 records where rewards have been given but the line item table does not have those receipt_ids

SELECT 'user_table ids' AS u_ids, COUNT(distinct user_id) AS count FROM users
WHERE user_id NOT IN (SELECT user_id FROM receipts_fact)
UNION
SELECT 'receipt_fact ids' AS u_ids, COUNT(distinct user_id) AS count FROM receipts_fact
WHERE user_id NOT IN (SELECT user_id FROM users);



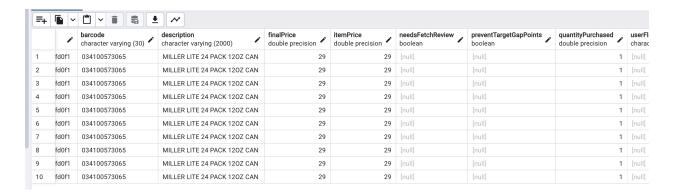
This is a clear violation for our data model which needs to be handled firstly. There are some user_ids present in receipts_fact table which are not in the users_dim table. Hence this leads to failure of referential integrity.

Also, let's analyze Main fact v/s line level details for one receipt_id -

select * from receipts fact where receipt id = '5ff1e1b60a7214ada100055c'

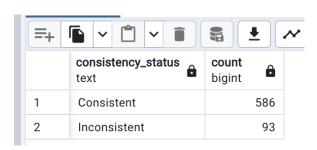


select * from receipts_line_item where receipt_id = '5ff1e1b60a7214ada100055c'



For this receipt_id it's consistent for the quantitypurchased (sum =10) and totalSpent = (price*10) = 290

Thus we can observe, that fact table is aggregated and line_item table contains granular/line level details for the receipt_id



group by consistency status

We observe that around 93 records are Inconsistent, w.r.t non matching between fact and line level fields.

On further analysis, I found that its due to Nulls as well as counts mismatch.

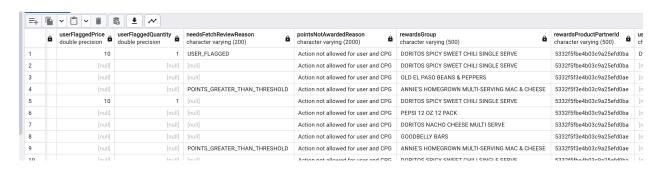
ALTER TABLE receipts_line_item

ADD CONSTRAINT pk_integration_id PRIMARY KEY (integration_id),

ADD CONSTRAINT fk_receipt_id FOREIGN KEY (receipt_id) REFERENCES receipts_fact(receipt_id);

Also,

select * from receipts_fact rf
inner join receipts_line_item rl on rf.receipt_id = rl.receipt_id
where rf."rewardsReceiptStatus" = 'FINISHED' and rl."pointsNotAwardedReason" is Not NULL



Need to understand context for pointsNotAwardedReason field.

select * from receipts_line_item where barcode = '4011'



How to handle around 170 records for ITEM NOT FOUND. Is it expected?