# Data Quality Issues

1. **Outdated data:**

select \* from rpts where datescanned >= (select DATEADD(MONTH, -12, current\_date));

**A screenshot of a computer

Description automatically generated**

**This query indicates that there is no data entry for datescanned in the last 1 year.**

1. **Missing data:**

select \* from brand\_flattened where brandcode='';

select \* from items where REWARDSRECEIPTITEMLIST\_BARCODE is null;

select \* from items where REWARDSRECEIPTITEMLIST\_FINALPRICE is null;

**A screenshot of a computer

Description automatically generated**

**There are missing values for brandcode in brand table, barcode and final price in item table which can lead to data integrity issues and lack of precision in the statistical analysis.**

1. **Irrelevant data:**

SELECT i.REWARDSRECEIPTITEMLIST\_BRANDCODE ,b.BRANDCODE

FROM items i

LEFT JOIN brand\_flattened b ON i.REWARDSRECEIPTITEMLIST\_BRANDCODE = b.BRANDCODE

WHERE b.BRANDCODE IS NULL;

**A screenshot of a computer

Description automatically generated**

**This indicates that some brand codes from the receipts do not have corresponding matches in the brand table which causes data discrepancy and thus is difficult to link both tables.**

Another data quality issue to look at:

WITH NullCount AS (

SELECT COUNT(\*) AS nc

FROM items

WHERE REWARDSRECEIPTITEMLIST\_FINALPRICE IS NULL

),

FullCount AS (

SELECT COUNT(\*) AS fc

FROM items

)

SELECT (NullCount.nc \* 100.0 / FullCount.fc) AS PercentageNull

FROM NullCount, FullCount;

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

**As we can see above, 40% of the data for final price is missing. Similarly, 82% of discount fields, 50% of barcode are missing which could effect the rewards system.**