CLOSE ENDED 1

Q:What are the top 5 brands by receipts scanned among users 21 and over?

-- a view of users who are 21 and over

WITH user\_above21 AS (

SELECT \*

FROM USER\_TAKEHOME

WHERE ( (strftime('%Y', 'now') - strftime('%Y', birth\_date)) > 21

        AND

            (strftime('%m-%d', 'now') < strftime('%m-%d', birth\_date))

          )

),

--a view of the transactions corresponding to the users 21 and over

transactions\_above21 as (

SELECT \*

FROM TRANSACTION\_TAKEHOME

WHERE user\_id IN (SELECT id FROM user\_above21)

),

–count the brands based on receipts scanned

SELECT p.BRAND, COUNT(t.receipt\_id) AS Total

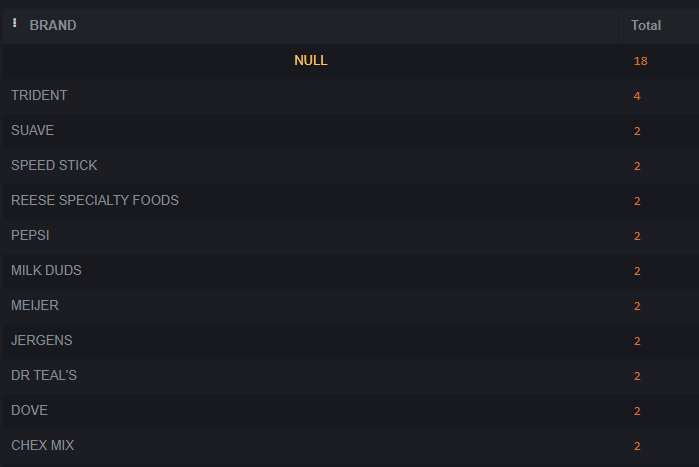
FROM transactions\_above21 t LEFT JOIN PRODUCTS\_TAKEHOME p on t.BARCODE = p.BARCODE

WHERE t.BARCODE IS NOT NULL

GROUP BY p.BRAND

ORDER BY 2 DESC

LIMIT 5;

– as you can see null is the highest which means that the brand name doesn’t exist, a data quality issue