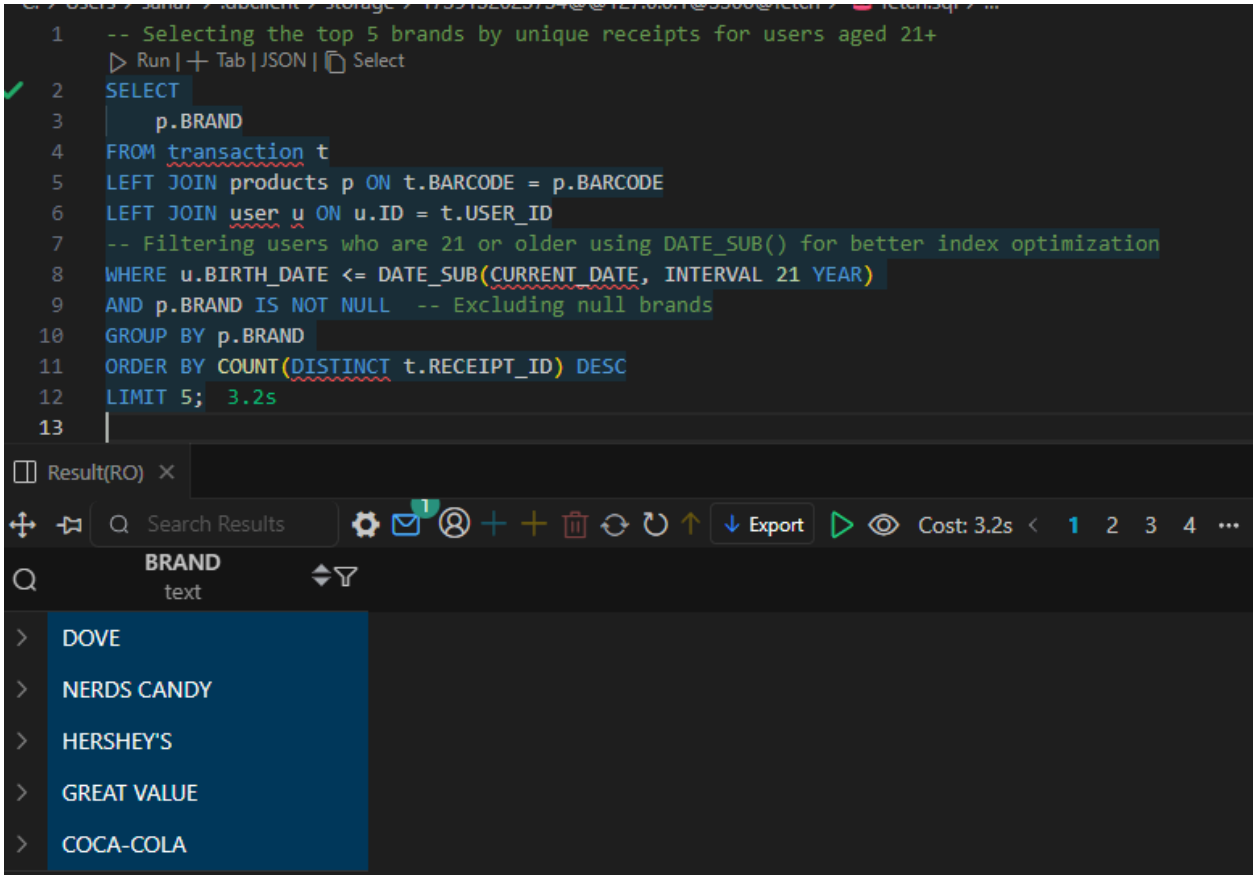
Section 2 - SQL Queries

Query for - What are the top 5 brands by receipts scanned among users 21 and over?

```sql
-- Selecting the top 5 brands by unique receipts for users aged 21+
SELECT
   p.BRAND
FROM transaction t
LEFT JOIN products p ON t.BARCODE = p.BARCODE
LEFT JOIN user u ON u.ID = t.USER_ID
-- Filtering users who are 21 or older using DATE_SUB() for better index optimization
WHERE u.BIRTH_DATE <= DATE_SUB(CURRENT_DATE, INTERVAL 21 YEAR)
AND p.BRAND IS NOT NULL  -- Excluding null brands
GROUP BY p.BRAND
ORDER BY COUNT(DISTINCT t.RECEIPT_ID) DESC
LIMIT 5;
```

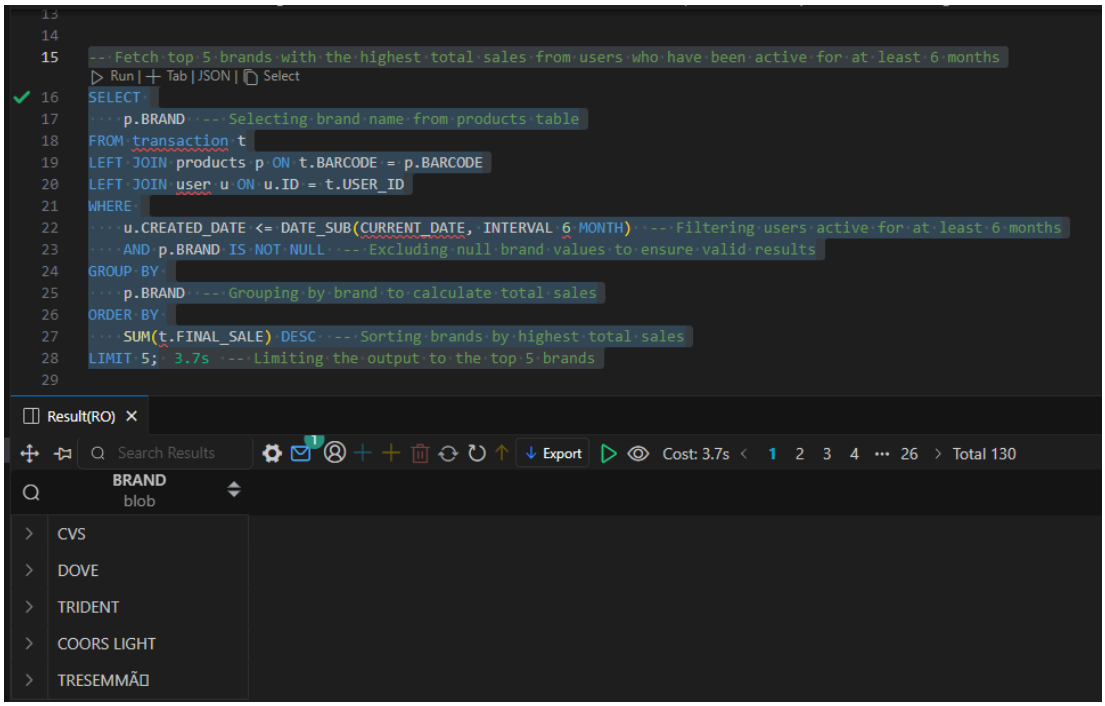
Query + Output



Query for - What are the top 5 brands by sales among users that have had their account for at least six months?

```sql
-- Fetch top 5 brands with the highest total sales from users who have been active for at least 6 months
SELECT
    p.BRAND  -- Selecting brand name from products table
FROM transaction t
LEFT JOIN products p ON t.BARCODE = p.BARCODE
LEFT JOIN user u ON u.ID = t.USER_ID
WHERE
    u.CREATED_DATE <= DATE_SUB(CURRENT_DATE, INTERVAL 6 MONTH)  -- Filtering users active for at least 6 months
    AND p.BRAND IS NOT NULL  -- Excluding null brand values to ensure valid results
GROUP BY
    p.BRAND  -- Grouping by brand to calculate total sales
ORDER BY
    SUM(t.FINAL_SALE) DESC  -- Sorting brands by highest total sales
LIMIT 5;  -- Limiting the output to the top 5 brands
```

Query + Output

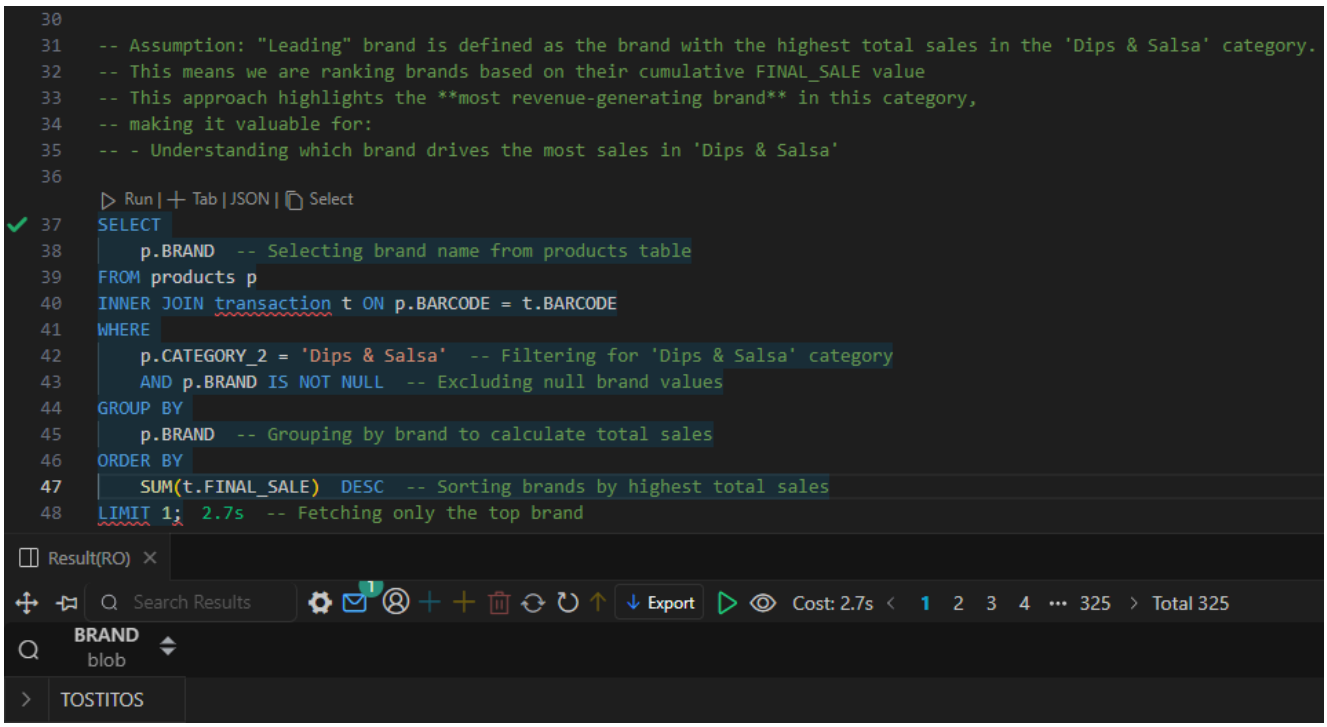


Question - Which is the leading brand in the Dips & Salsa category?

```sql
-- Assumption: "Leading" brand is defined as the brand with the highest total sales in the 'Dips & Salsa' category.
-- This means we are ranking brands based on their cumulative FINAL_SALE value
-- This approach highlights the **most revenue-generating brand** in this category,
-- making it valuable for:
-- - Understanding which brand drives the most sales in 'Dips & Salsa'

SELECT
    p.BRAND  -- Selecting brand name from products table
FROM products p
INNER JOIN transaction t ON p.BARCODE = t.BARCODE
WHERE
    p.CATEGORY_2 = 'Dips & Salsa'  -- Filtering for 'Dips & Salsa' category
    AND p.BRAND IS NOT NULL  -- Excluding null brand values
GROUP BY
    p.BRAND  -- Grouping by brand to calculate total sales
ORDER BY
    SUM(t.FINAL_SALE)  DESC  -- Sorting brands by highest total sales
LIMIT 1;  -- Fetching only the top brand
```

Query + Output



Query - At what percent has Fetch grown year over year?

```sql
-- Assumption: Fetch's year-over-year growth is measured by the number of new users onboarded each year.
-- This insight demonstrates the platform's expanding reach and increasing consumer participation,
-- making it a valuable tool for CPG brands and retailers to analyze shopping trends, loyalty patterns, and purchase behaviors.

WITH sum_of_users_per_year AS (
  -- Calculate the number of unique users onboarded per year
  SELECT
    YEAR(created_date) AS current_year,
    COUNT(DISTINCT id) AS count_of_unique_users,
    LAG(COUNT(DISTINCT id)) OVER (ORDER BY YEAR(created_date)) AS count_of_unique_users_last_year
  FROM user
  GROUP BY YEAR(created_date)
)
-- Calculate the YoY growth percentage
SELECT
  current_year,
  IFNULL(((count_of_unique_users - count_of_unique_users_last_year) * 100 / count_of_unique_users_last_year),0) AS yoy_growth_percentage
FROM sum_of_users_per_year;
```

Query+output