

## Analysis by SQL Query

Closed-ended questions:

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

SQL Query:

```
SELECT p.BRAND, COUNT(t.RECEIPT_ID) AS receipt_count
FROM Fetch.transaction t
JOIN Fetch.user u ON t.USER_ID = u.ID
JOIN Fetch.products p ON t.BARCODE = p.BARCODE
WHERE TIMESTAMPDIFF(YEAR, u.BIRTH_DATE, CURDATE()) >= 21
GROUP BY p.BRAND
ORDER BY receipt_count DESC
LIMIT 5;
```

Output: COCA-COLA, ANNIE'S HOMEGROWN GROCERY, DOVE, BAREFOOT, ORIBE

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

SQL Query:

```
SELECT p.BRAND, SUM(t.FINAL_SALE) AS total_sales
FROM Fetch.transaction t
JOIN Fetch.user u ON t.USER_ID = u.ID
JOIN Fetch.products p ON t.BARCODE = p.BARCODE
WHERE u.CREATED_DATE <= DATE_SUB(CURDATE(), INTERVAL 6 MONTH)
GROUP BY p.BRAND
ORDER BY total_sales DESC
LIMIT 5;
```

Output: COCA-COLA, ANNIE'S HOMEGROWN GROCERY, DOVE, BAREFOOT, ORIBE

What is the percentage of sales in the Health & Wellness category by generation?

SQL Query:

```
SELECT
CASE
    WHEN YEAR(u.BIRTH_DATE) BETWEEN 2013 AND 2025 THEN 'Generation Alpha'
    WHEN YEAR(u.BIRTH_DATE) BETWEEN 1997 AND 2012 THEN 'Generation Z'
    WHEN YEAR(u.BIRTH_DATE) BETWEEN 1981 AND 1996 THEN 'Millennials/Generation Y'
    WHEN YEAR(u.BIRTH_DATE) BETWEEN 1965 AND 1980 THEN 'Generation X'
    WHEN YEAR(u.BIRTH_DATE) BETWEEN 1946 AND 1964 THEN 'Baby Boomers'
    WHEN YEAR(u.BIRTH_DATE) BETWEEN 1928 AND 1945 THEN 'Silent Generation'
    WHEN YEAR(u.BIRTH_DATE) BETWEEN 1901 AND 1927 THEN 'Greatest/G.I. Generation'
    ELSE 'Unknown'
```

```

        END AS generation,
        SUM(t.FINAL_SALE) / (SELECT SUM(FINAL_SALE) FROM Fetch.transaction) * 100 AS
percentage_sales
FROM Fetch.transaction t
JOIN Fetch.user u ON t.USER_ID = u.ID
JOIN Fetch.products p ON t.BARCODE = p.BARCODE
WHERE p.CATEGORY_1 = 'Health & Wellness'
GROUP BY generation
ORDER BY percentage_sales DESC;

```

Output:

GENERATION	
Baby Boomers	0.082372
Generation X	0.066505
Millennials/Generation Y	0.126852

Reference Table to Group By Generation:

- [https://upload.wikimedia.org/wikipedia/commons/3/3e/Generation\\_timeline.svg](https://upload.wikimedia.org/wikipedia/commons/3/3e/Generation_timeline.svg)

Open-ended questions:

Who are Fetch's power users?

Power users are defined as those who have scanned the highest number of receipts.

SQL Query:

```

SELECT u.ID, COUNT(t.RECEIPT_ID) AS total_receipts
FROM Fetch.transaction t
JOIN Fetch.user u ON t.USER_ID = u.ID
GROUP BY u.ID
ORDER BY total_receipts DESC
LIMIT 10;

```

Output:

64e62de5ca929250373e6cf5	80502
604278958fe03212b47e657b	72452
66390784b7b24d45d93a0e6a	40252
64023fa080552327896edb23	40252
63ae0dc29f3fc9c7546ef080	32204
5d8661a736d69e65e99233af	32202
61aea787e9b3d75037b5ea45	24156
63d97d69b425eb11a4709d56	24154
634aee03305e373439460ac3	24154
6456eac19f7c516a13f471f4	24154

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

SQL Query:

```

SELECT p.BRAND, SUM(t.FINAL_SALE) AS total_sales
FROM Fetch.transaction t
JOIN Fetch.products p ON t.BARCODE = p.BARCODE

```

```
WHERE p.CATEGORY_2 = 'Dips & Salsa'  
GROUP BY p.BRAND  
ORDER BY total_sales DESC  
LIMIT 1;
```

Output: TOSTITOS

At what percent has Fetch grown year over year?

- Year-over-year analysis is not possible.
- The Exploring Data Analysis report confirms that the transaction table only contains records from 2024, making year-over-year growth calculations infeasible.

SQL Query:

```
WITH yearly_sales AS (  
  SELECT YEAR(PURCHASE_DATE) AS year, SUM(FINAL_SALE) AS total_sales  
  FROM Fetch.transaction  
  GROUP BY year  
)  
SELECT  
  y1.year AS current_year,  
  y2.year AS previous_year,  
  (y1.total_sales - y2.total_sales) / y2.total_sales * 100 AS growth_percentage  
FROM yearly_sales y1  
JOIN yearly_sales y2 ON y1.year = y2.year + 1  
ORDER BY current_year DESC;
```

Output: