Close-Ended Questions

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

```
SELECT
    p.BRAND
  , COUNT(DISTINCT t.RECEIPT_ID) as num_receipts
FROM product as p
JOIN "transaction" as t
ON p.BARCODE = t.BARCODE
JOIN user as u
ON t.USER_ID = u.ID
WHERE u.BIRTH_DATE <=  DATE('now', '-21 years') -- Users aged 21 and over
GROUP BY 1
ORDER BY 2 DESC
LIMIT 5
```

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

```
SELECT
  p.BRAND
  , SUM(t.FINAL_SALE) AS total_sales
FROM "transaction" as t
JOIN product as p
ON t.BARCODE = p.BARCODE
JOIN user as u
ON t.USER_ID = u.ID
WHERE DATE('now', '-6 months') >= u.CREATED_DATE -- Users that have had
their account for at least six months
GROUP BY 1
ORDER BY 2 DESC
LIMIT 5
```

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

```
WITH cte AS (
SELECT
    u.ID
  , u.BIRTH_DATE
  , CASE
        WHEN CAST(strftime('%Y', BIRTH_DATE) AS INTEGER) BETWEEN 1928 AND
1945 THEN 'Silent Generation'
        WHEN CAST(strftime('%Y', BIRTH_DATE) AS INTEGER) BETWEEN 1946 AND
1964 THEN 'Baby Boomers'
        WHEN CAST(strftime('%Y', BIRTH_DATE) AS INTEGER) BETWEEN 1965 AND
```

```
1980 THEN 'Gen X'
        WHEN CAST(strftime('%Y', BIRTH_DATE) AS INTEGER) BETWEEN 1981 AND
1996 THEN 'Millennials'
        WHEN CAST(strftime('%Y', BIRTH_DATE) AS INTEGER) BETWEEN 1997 AND
2012 THEN 'Gen Z'
        WHEN CAST(strftime('%Y', BIRTH_DATE) AS INTEGER) >= 2013 THEN 'Gen
Alpha'
        ELSE 'Unknown'
    END AS generation
  , t.FINAL_SALE
FROM "transaction" as t
JOIN product as p
ON t.BARCODE = p.BARCODE
JOIN user as u
ON t.USER_ID = u.ID
WHERE p.CATEGORY_1 = 'Health & Wellness'
)
SELECT
    c.generation
  , ROUND(SUM(c.FINAL_SALE)*100 / (SELECT SUM(FINAL_SALE) FROM CTE), 2) AS
health_wellness_percentage
FROM cte as c
GROUP BY 1
ORDER BY 2 DESC
```

Open-Ended Questions

1. Who are Fetch's power users?

Typically, a power user is someone who engages with the product more intensely than the average user. This could mean they make more transactions, use more features, or are more active over a certain period.

To define power users, we need to establish criteria based on user activity and engagement. Potential metrics could include:

- **Frequency of Transactions**: Number of transactions a user has made.
- **Total Spending**: The cumulative amount spent by a user.
- **Recency of Activity**: How recently a user has been active.
- **Product Diversity**: Number of different products a user has purchased.

For this case study, let's define a power user as someone who:

- Has spent more than $500 in total.

```
SELECT
    u.ID
  , SUM(t.FINAL_SALE) as total_spend
FROM user as u
```

```
JOIN "transaction" as t
ON u.ID = t.USER_ID
GROUP BY 1
HAVING SUM(t.FINAL_SALE) > 500
ORDER BY 2 DESC
```

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

The **leading brand** can be defined based on different metrics such as:

- **Total Sales Revenue**: The brand with the highest total sales amount in the "Dips & Salsa" category.
- **Number of Transactions**: The brand with the most transactions in the "Dips & Salsa" category.
- **Number of Unique Customers**: The brand with the most unique users purchasing their products in the category.

For this case study, let's define the leading brand as the one with the highest total sales revenue in the "Dips & Salsa" category.

```
WITH CTE AS (
SELECT
  p.BRAND
  , COUNT(DISTINCT t.USER_ID) AS unique_customers
  , DENSE_RANK() OVER(ORDER BY SUM(t.FINAL_SALE)) AS rnk
FROM "transaction" as t
JOIN product as p
ON t.BARCODE = p.BARCODE
WHERE p.CATEGORY_2 = 'Dips & Salsa'
GROUP BY 1
)

SELECT
  cte.BRAND
  , cte.unique_customers
FROM CTE
WHERE rnk=1
```

3. At what percent has Fetch grown year over year?

To calculate Fetch's year-over-year (YoY) growth, we need to determine the percentage increase or decrease in a specific metric over two consecutive years. Common metrics for growth analysis include:

- **Total Revenue**: The sum of all transaction amounts.
- **Number of Transactions**: The total count of transactions.
- **Number of Active Users**: The count of unique users making transactions.

For this case study, assuming the current year is 2024 and the previous year is 2023; let's focus on total revenue as the metric for growth calculation.

Year-over-year revenue growth is calculated using the formula:

$$YoY\ Revenue\ Growth\ =\ \frac{Total\ Revenue\ in\ Current\ \ Year\ -\ Total\ Revenue\ in\ Previous\ Year}{Total\ Revenue\ in\ Previous\ Year}\ *\ 100$$

```
WITH cte AS (
SELECT
    strftime('%Y', t.SCAN_DATE) AS transaction_year
  , SUM(t.FINAL_SALE) AS total_revenue
FROM "transaction" t
WHERE strftime('%Y', t.SCAN_DATE) IN (2023, 2024)
GROUP BY 1
)
SELECT
    MAX(CASE WHEN transaction_year = 2024 THEN total_revenue END) AS
current_year_revenue
  , MAX(CASE WHEN transaction_year = 2023 THEN total_revenue END) AS
previous_year_revenue
  , (MAX(CASE WHEN transaction_year = 2024 THEN total_revenue END) -
MAX(CASE WHEN transaction_year = 2023 THEN total_revenue END)) / MAX(CASE
WHEN transaction_year = 2023 THEN total_revenue END) * 100 AS
yoy_growth_percent
FROM cte
```