

Second: SQL Queries

Close-ended questions:

Q1 What are the top 5 brands by receipts scanned among users 21 and over?

```
SELECT TOP 5 p.BRAND, COUNT(t.RECEIPT_ID) as total_receipts
FROM Users as u
JOIN Transactions as t
    ON u.ID = t.USER_ID
JOIN Products as p
    ON p.BARCODE = t.BARCODE
WHERE u.AGE >= 21
GROUP BY p.BRAND
ORDER BY total_receipts DESC;
```

	BRAND	total_receipts
1	CORICIDIN® HBP	5
2	NERDS CANDY	3
3	DOVE	3
4	SOUR PATCH KIDS	2
5	GREAT VALUE	2

Q2 What are the top 5 brands by sales among users that have had their account for at least six months?

```
SELECT TOP 5 p.BRAND, SUM(t.FINAL_SALE) as sales
FROM Users as u
JOIN Transactions as t
    ON u.ID = t.USER_ID
JOIN Products as p
    ON p.BARCODE = t.BARCODE
WHERE
    DATEDIFF(MONTH, u.CREATED_DATE, GETDATE()) >= 6
    AND t.PURCHASE_DATE <= GETDATE()
GROUP BY p.BRAND
ORDER BY sales DESC;
```

	BRAND	sales
1	CVS	72
2	DOVE	31.0425004959106
3	TRIDENT	23.3600006103516
4	CORICIDIN® HBP	21.3228063583374
5	COORS LIGHT	17.4799995422363

Q3 What is the percentage of sales in the Health & Wellness category by generation?

```
WITH UserGeneration AS (
    SELECT
        u.ID,
        CASE
            WHEN u.AGE BETWEEN 0 AND 24 THEN 'Gen Z'
            WHEN u.AGE BETWEEN 25 AND 40 THEN 'Millennials'
            WHEN u.AGE BETWEEN 41 AND 56 THEN 'Gen X'
            WHEN u.AGE BETWEEN 57 AND 75 THEN 'Baby Boomers'
            ELSE 'Silent'
        END AS generation
    FROM Users as u)
SELECT
    ug.generation,
    SUM(CASE WHEN p.CATEGORY_1 = 'Health & Wellness' THEN t.FINAL_SALE
    ELSE 0 END) AS health_and_wellness_sales,
    SUM(t.FINAL_SALE) AS total_sales,
    (SUM(CASE WHEN p.CATEGORY_1 = 'Health & Wellness' THEN t.FINAL_SALE
    ELSE 0 END) / SUM(t.FINAL_SALE)) * 100 AS percentage_of_sales
FROM UserGeneration as ug
JOIN Transactions as t
    ON ug.ID = t.USER_ID
JOIN Products as p
    ON p.BARCODE = t.BARCODE
GROUP BY ug.generation;
```

	generation	health_and_wellness_sales	total_sales	percentage_of_sales
1	Gen X	51.1424999237061	118.229306042194	43.2570414525262
2	Millennials	33.2424070835114	77.485964179039	42.901198217914
3	Silent	3.59333324432373	10.3033332824707	34.8754441481297
4	Baby Boomers	105.789458632469	183.65758395195	57.6014648325912

Open-ended questions:

Q1 Who are Fetch's power users?

Assumptions:

1. Long Account History - Users with accounts older than 2 years.
2. They purchase almost every month or frequently.
3. They always purchase the most expensive brand.
4. They make the maximum amount of sales(High-spending, Users in the top 10% of sales).

```
WITH UserPurchaseStats AS (  
    -- Calculate total spending, purchase frequency, and oldest account  
    age  
    SELECT  
        t.USER_ID,  
        COUNT(t.RECEIPT_ID) AS total_purchases, -- Count total number  
of purchases  
        COUNT(DISTINCT FORMAT(t.PURCHASE_DATE, 'yyyy-MM')) as  
purchase_frequency, -- Count unique months with purchases  
        SUM(t.FINAL_SALE) as total_sales, -- Total amount spent by  
users  
        DATEDIFF(YEAR, u.CREATED_DATE, GETDATE()) as account_age --  
Account age in years  
    FROM  
        Transactions as t  
    INNER JOIN  
        Users as u ON t.USER_ID = u.ID  
    GROUP BY  
        t.user_id, u.CREATED_DATE  
)  
TopSpenders AS (  
    -- Identify users in the top 10% of total sales  
    SELECT  
        USER_ID  
    FROM  
        UserPurchaseStats  
    WHERE  
        total_sales >= (SELECT PERCENTILE_CONT(0.90) WITHIN GROUP  
(ORDER BY total_sales) OVER ())
```

```

),
TopBrands AS (
    -- Calculate the price dynamically
    SELECT TOP 5 p.BRAND,
               AVG(CASE
                    WHEN t.FINAL_QUANTITY = 0 THEN 0
                    ELSE t.FINAL_SALE/ t.FINAL_QUANTITY
                END) AS avg_price
    FROM Products AS p
    INNER JOIN Transactions AS t ON p.BARCODE = t.BARCODE
    GROUP BY p.BRAND
    ORDER BY avg_price DESC
),
LoyalPremiumBuyers AS (
    -- Identify users who frequently buy premium brands
    SELECT
        t.USER_ID
    FROM
        Transactions AS t
    INNER JOIN
        Products AS p ON t.BARCODE = p.BARCODE
    WHERE
        p.BRAND IN (SELECT tb.BRAND FROM TopBrands AS tb)
    GROUP BY
        t.USER_ID
    HAVING
        COUNT(DISTINCT p.BRAND) <= 2 -- Ensure brand Loyalty by
limiting to 1-2 brands
)
SELECT
    ups.USER_ID,
    ups.account_age,
    ups.total_purchases,
    ups.purchase_frequency,
    ups.total_sales
FROM
    UserPurchaseStats AS ups
INNER JOIN
    TopSpenders ts ON ups.user_id = ts.user_id
INNER JOIN
    LoyalPremiumBuyers lpb ON ups.user_id = lpb.user_id
WHERE
    ups.account_age >= 2

```

```

    AND ups.purchase_frequency >= 2
ORDER BY
    ups.total_sales DESC;

```

	USER_ID	account_age	total_purchases	purchase_frequency	total_sales
1	62ffec490d9dbaff18c0a999	3	3	3	26.1400005817413
2	61a58ac49c135b462ccddd1c	4	3	2	14.9816663265228
3	632fc9dc0c625b72ae991f83	3	2	2	14.4122223854065
4	646bdaa67a342372c857b958	2	3	2	13.3824999332428
5	62c09104baa38d1a1f6c260e	3	3	3	11.3087499141693
6	63c2d13139c79dcbdd563f61	2	2	2	11.039999961853
7	5c366bf06d9819129dfa1118	6	3	3	9.90142834186554
8	5f21e60446f11314a16015de	5	2	2	8.98000001907349
9	65044dc5fe41d365c2ed7d71	2	2	2	7.86812520027161
10	63c8294d39c79dcbdd5c1e4f	2	2	2	7.8600001335144
11	5ca54049adfc4140b54e1bf	6	2	2	6.53280639648438
12	5c6c2998b06c0045a4fddc74	6	2	2	5.46000003814697
13	61fd8c73ac4e5e6b50b4ca01	3	2	2	4.46803033351898
14	61be248f43595e62de653970	4	2	2	2.99000000953674

Q2 Which is the leading brand in the Dips & Salsa category?

```

SELECT TOP 5 p.BRAND, SUM(t.FINAL_SALE) as total_sales
FROM Transactions as t
INNER JOIN Products as p
    ON t.BARCODE = p.BARCODE
WHERE p.CATEGORY_2 = 'Dips & Salsa'
GROUP BY p.BRAND
ORDER BY total_sales DESC;

```

	BRAND	total_sales
1	TOSTITOS	169.00334405899
2	NULL	168.531247615814
3	PACE	85.5975009202957
4	FRITOS	67.0262498855591
5	HELUVA GOOD!	53.0027282238007

Q3 At what percent has Fetch grown year over year?

```
WITH YearlySales AS (  
    SELECT  
        YEAR(t.PURCHASE_DATE) AS sales_year,  
        SUM(t.FINAL_SALE) AS annual_sales  
    FROM Transactions as t  
    GROUP BY YEAR(t.PURCHASE_DATE)  
) ,  
YoYGrowth AS (  
    SELECT  
        a.sales_year AS current_year,  
        a.annual_sales AS current_sales,  
        b.annual_sales AS previous_sales,  
        ((a.annual_sales - b.annual_sales) * 100.0/  
ISNULL(b.annual_sales, 0)) AS yoy_growth_percent  
    FROM YearlySales a  
    LEFT JOIN YearlySales b ON a.sales_year = b.sales_year + 1  
)  
SELECT * FROM YoYGrowth  
ORDER BY current_year DESC;
```

	current_year	current_sales	previous_sales	yoy_growth_percent
1	2024	100629.890433621	NULL	NULL

Third: communicate with stakeholders

Subject: Data Quality Issues, Consumer Trends & Recommendations – Need Clarification

Hi Client Name,

During our analysis of purchase trends, we identified a few **data quality concerns** that need clarification to ensure accurate insights:

- **Product Categorization & Manufacturer Data** – Missing values in the manufacturer column and inconsistent product categorization could impact reporting.
- **Final Quantity & Sales Discrepancies** – Many records show **zero Final Quantity** while having **non-zero Final Sales**, which raises concerns about incorrect data entry. Should we exclude these records or assume a default quantity?
- **Basket Size Limitation** – In the transaction data, each receipt is associated with only two barcodes, making the basket size too small for meaningful analysis. A more

comprehensive dataset (with basket sizes >2) would help us track shopping behavior trends more effectively.

Key Consumer Trends & Insights:

Health & Wellness vs. Snacks – **Snacks** is the most purchased category across all generations, with Boomers (42-57 years old) driving maximum purchases. However, Health & Wellness leads in overall sales revenue, indicating a possible shift towards healthier lifestyles. **CVS** is the leading brand across all categories.

Shopping Patterns:

- Saturday is the busiest shopping day.
- August records the highest sales volume.
- **Nerds Candy & Trident** are frequently purchased on repeat.
- **Power Users Identified:** We found 14 top power users who have accounts older than 2 years, purchase frequently, prefer premium brands, and are high spenders.
- 4 Users are purchasing both Snacks & Health & Wellness.

Recommendations & Next Steps:

- **Targeted Discounts & Fetch Rewards** – To encourage healthier choices, we recommend sending discounts and Fetch rewards to users for CVS and Health & Wellness products, especially for users transitioning from Snacks to Health & Wellness.
- **Handling Duplicate Transactions & Missing Values** – Should we remove duplicate transactions, or are they expected? How should we treat zero Final Quantity with non-zero Final Sales values?

Clarification on Basket Size & Data Accuracy – Can we access data with larger basket sizes (>2 items per receipt) to analyze true purchasing behavior?

- **Access to Marketing Campaign Data** – This will help determine whether recent growth trends are organic or driven by past promotions.(Need Historical data)

Please share your thoughts on these issues so we can refine our insights and ensure data accuracy.

Thanks,

Chelsea.