FETCH SQL QUERIES

Closed-ended questions:

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

Assumption:

In this query I calculate the age of Users in where clause and filter out data to only users age on or above 21. Since we wanted to check the result for all the brands, I took Products table to be base table and left joined transactions and users to get all the rows/brands from products table. Since they asked to get the result by receipts, I calculated a count of distinct receipts scanned for each brand and then used TOP function to get the TOP 5 brands by receipts.

Query:

```
SELECT  TOP 5                    #to get top 5 brands
A.BRAND,
COUNT( DISTINCT B.RECEIPT_ID) AS COUNT_OF_RECEIPTS
FROM
 [dbo].[PRODUCTS_TAKEHOME] A
LEFT JOIN [dbo].[TRANSACTION_TAKEHOME] B
ON A.BARCODE=B.BARCODE
LEFT JOIN [dbo].[USER_TAKEHOME] C
ON B.USER_ID=C.ID
WHERE DATEDIFF(YEAR,C.BIRTH_DATE,GETDATE())>=21    #to get users of age>=21
AND A.BRAND IS NOT NULL  #to filter out sales where brand name is null
GROUP BY A.BRAND             #to aggregate result by each  brand
ORDER BY COUNT_OF_RECEIPTS DESC;
```

Result:

BRAND	COUNT_OF_RECEIPTS
DOVE	3
NERDS CANDY	3
COCO-COLA	2
GREAT VALUE	2
HERSHEY'S	2

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

Assumption:

In this query, to calculate if the users have had their account for 6 months, I calculated the difference between created date (assumed that's when users account was created at Fetch) and Today's date and checked it is at least 6 months. Since we wanted to check the result for all the brands, I took Products table to be base table and left joined transactions and users to get all the rows/brands from products table. Since they asked to get the result by sales, I calculated a sum of final sales for each brand and then used TOP function to get the TOP 5 brands by receipts.

Here I assumed that Final quantity in Transactions table is weight/volume of the product because it is given as whole number, that is why I did not take the total sales to be product of final quantity and final sales. I just considered Final Sales as the total sales.

Query:

```
SELECT TOP 5                    #to get top 5 brands
A.BRAND,
SUM(B.FINAL_SALE) AS FINAL_SALES
FROM
[dbo].[PRODUCTS_TAKEHOME] A
LEFT JOIN [dbo].[TRANSACTION_TAKEHOME] B
ON A.BARCODE =B.BARCODE
LEFT JOIN [dbo].[USER_TAKEHOME] C
ON B.USER_ID=C.ID
WHERE DATEDIFF(MONTH,C.CREATED_DATE,GETDATE())>=6   #to get users whose account is created >= 6 months
 AND A. BRAND IS NOT NULL          #to filter out sales where brand name is null
 GROUP BY A.BRAND                  #to aggregate result by each  brand
 ORDER BY FINAL_SALES DESC;
```

Result:

BRAND	FINAL_SALES
CVS	72
TRIDENT	46.72
DOVE	42.88
COORS LIGHT	34.96
QUAKER	16.6

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

Assumption:

In this query by generation, I assumed it was user generation, that means we want the percentage of sales in that given category per generation. So, I added a calculation for generation in CUSTOMER_GENERATIONS cte. For the final outer query, again I used Products as a base table to be left joined with transactions and users to see how each generation is having sales in that category.

Query:

```
WITH TOTAL_SALES AS (                #CTE to generate Total Sales for Health & Wellness Category
SELECT
 A. CATEGORY_1 AS CATEGORY,
 SUM (B. FINAL_SALE) AS TOTAL_SALES
FROM
[dbo]. [PRODUCTS_TAKEHOME] A
LEFT JOIN [dbo]. [TRANSACTION_TAKEHOME] B
ON A. BARCODE = B. BARCODE
WHERE A. CATEGORY_1 LIKE '%Health & Wellness%'    #to filter out Health & Wellness Category
GROUP BY A. CATEGORY_1
 ),

CUSTOMER_GENERATIONS AS (  #CTE to check which generation does each user belong to based on their Birth date
 SELECT
 ID,
 CASE
    WHEN YEAR(BIRTH_DATE) <= 1964 THEN 'Other'
    WHEN YEAR(BIRTH_DATE) BETWEEN 1965 AND 1980 THEN 'Generation X'
    WHEN YEAR(BIRTH_DATE) BETWEEN 1981 AND 1996 THEN 'Millennials'
    WHEN YEAR(BIRTH_DATE) >= 1997 THEN 'Generation Z'
    ELSE 'Other'
  END AS GENERATION
 FROM
 [dbo]. [USER_TAKEHOME]
)

SELECT
 A. CATEGORY_1,
 Gen. GENERATION,
 SUM (B. FINAL_SALE) AS FINAL_SALE,
 C.TOTAL_SALES,
```

```
 (SUM (B. FINAL_SALE) / C. TOTAL_SALES) * 100 AS Percentage_Sales  #dividing sales for each generation by total sales of the category
FROM
[dbo]. [PRODUCTS_TAKEHOME] A
LEFT JOIN [dbo]. [TRANSACTION_TAKEHOME] B
ON A. BARCODE = B. BARCODE
LEFT JOIN TOTAL_SALES C
ON A. CATEGORY_1 = C. CATEGORY
LEFT JOIN CUSTOMER_GENERATIONS Gen
ON B.USER_ID = Gen.ID
WHERE A. CATEGORY_1 LIKE '%Health & Wellness%'    #to filter out Health & Wellness Category
GROUP BY
  A. CATEGORY_1,
  Gen.Generation,
  C.TOTAL_SALES
ORDER BY PERCENTAGE_SALES DESC;
```

Result:

CATEGORY_1	GENERATION	FINAL_SALES	TOTAL_SALES	PERCENTAGE_SALES
Health & Wellness	NULL	41145.3800000007	41335.0400000008	99.5411641067723
Health & Wellness	Others	89.03	41335.0400000008	0.215386267921837
Health & Wellness	Millennials	59.13	41335.0400000008	0.143050545009752
Health & Wellness	Generation X	41.5	41335.0400000008	0.100399080296038

Open-ended questions:

1) Who are Fetch's power users?

Assumption:

In this question I assumed that Fetch's power users means users who scan more number of receipts and make more sales. But since very less percent of users are present in transactions (~!%) so the values are fluctuated and the total receipts are relatively less but Other wise we can set a threshold for the users for total scans and total sales for them to become power users.

Query:

```
SELECT A.ID,
```

```
COUNT (DISTINCT B. RECEIPT_ID) AS TOTAL_SCANS ,
SUM(FINAL_SALE) AS TOTAL_SALES
FROM [DBO]. [USER_TAKEHOME] A
LEFT JOIN [dbo]. [TRANSACTION_TAKEHOME] B
ON A. ID=B.USER_ID
GROUP BY A.ID
ORDER BY TOTAL_SALES DESC;
```

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

Assumption:

In this query I assumed that the brand which has more number of receipts associated with it, that is which has more users buying the dips in it, gets to be the leading brand in that category so that we can give more offers or change the pricing accordingly. Here I have used Category_2 instead of Category_1 because the Category_1 for Dips & Salsa is Snacks.

Query:

```
SELECT TOP 1        #to get the topmost brand
A. BRAND,
COUNT (DISTINCT B. RECEIPT_ID) AS COUNT_OF_RECEIPTS
FROM
[dbo]. [PRODUCTS_TAKEHOME] A
LEFT JOIN [dbo]. [TRANSACTION_TAKEHOME] B
ON A. BARCODE=B.BARCODE
WHERE UPPER (A. CATEGORY_2) LIKE '%DIPS & SALSA%' #to filter out Dips & Salsa Category
AND A.BRAND IS NOT NULL              #to filter out sales where brand name is null
GROUP BY A. BRAND
ORDER BY COUNT_OF_RECEIPTS DESC;
```

Result:

BRAND	COUNT_OF_RECEIPTS
TOSTITOS	36

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

Assumption:

In this query I assumed that growth of fetch can be achieved by getting more users to interact with the app, so I have taken user created date (join date of user at Fetch) as base and counted how many users were there in each year and then calculated growth for each year when compared to previous years.

Query:

```
 WITH GROWTH AS (
 SELECT
YEAR(CREATED_DATE) as YEAR,
COUNT (DISTINCT ID) AS CURRENT_YEAR_USERS,
LAG (COUNT (DISTINCT ID)) OVER (ORDER BY YEAR(CREATED_DATE)) AS
PREVIOUS_YEAR_USERS     #lag function to get previous year enrolled users
FROM [dbo]. [USERS_TAKEHOME] A
 GROUP BY YEAR(CREATED_DATE)
 )
 SELECT
YEAR,
CURRENT_YEAR_USERS,
PREVIOUS_YEAR_USERS,
CASE
   WHEN PREVIOUS_YEAR_USERS IS NOT NULL AND PREVIOUS_YEAR_USERS ! = 0
   THEN
   CAST ((CURRENT_YEAR_USERS - PREVIOUS_YEAR_USERS) AS FLOAT) /
PREVIOUS_YEAR_USERS * 100
   ELSE NULL
END AS “YOY_GROWTH%” #calculation to get %Growth over previous year
FROM GROWTH;
```

Result:

YEAR	CURRENT_YEAR_USERS	PREVIOUS_YEAR_USERS	YOY_GROWTH%
2014	30	NULL	NULL
2015	51	30	70
2016	70	51	37.25
2017	644	70	820
2018	2168	644	236.65
2019	7093	2168	227.17
2020	16883	7093	138.02
2021	19159	16883	13.48
2022	26807	19159	39.92
2023	15464	26807	-42.31

2024	11631	15464	-24.79

I have conducted similar analysis for the above business questions in Power BI for easier understanding and attached below is the snippet of the results. I have added the following calculations for my analysis:

Purchase Day:

Date.DayOfWeek([PURCHASE_DATE]))

YOY% :

```
YoY% =
IF(
   ISFILTERED('USERS'[CREATED_DATE]),
   ERROR("Time intelligence quick measures can only be grouped or filtered by the Power BI-provided date hierarchy or primary date column."),
   VAR __PREV_YEAR =
      CALCULATE(
         COUNTA('USERS'[ID]),
         DATEADD('USERS'[CREATED_DATE].[Date], -1, YEAR)
      )
   RETURN
      DIVIDE(COUNTA('USERS'[ID]) - __PREV_YEAR, __PREV_YEAR)
)
```

Customer Age:

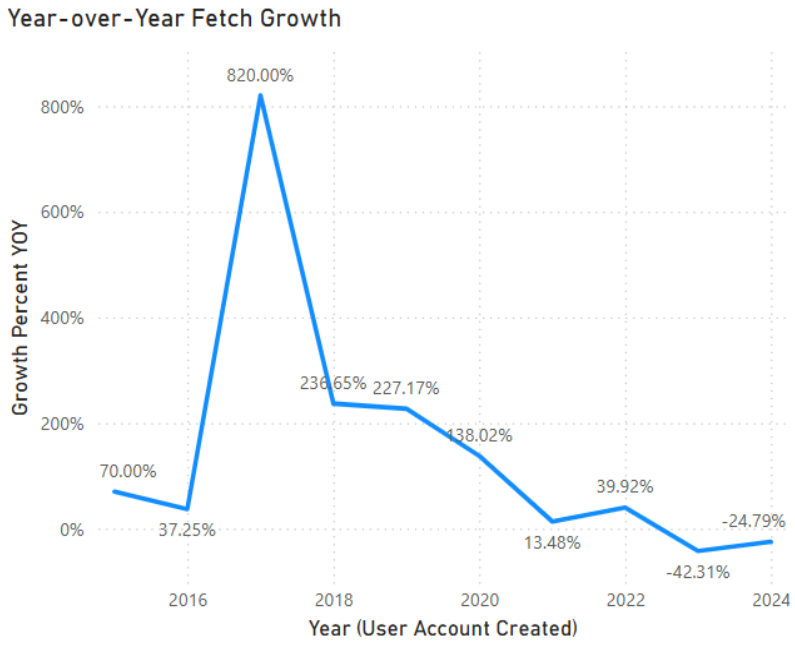
(Date.Year(DateTime.LocalNow()) - Date.Year([BIRTH_DATE])

Customer Account Age:

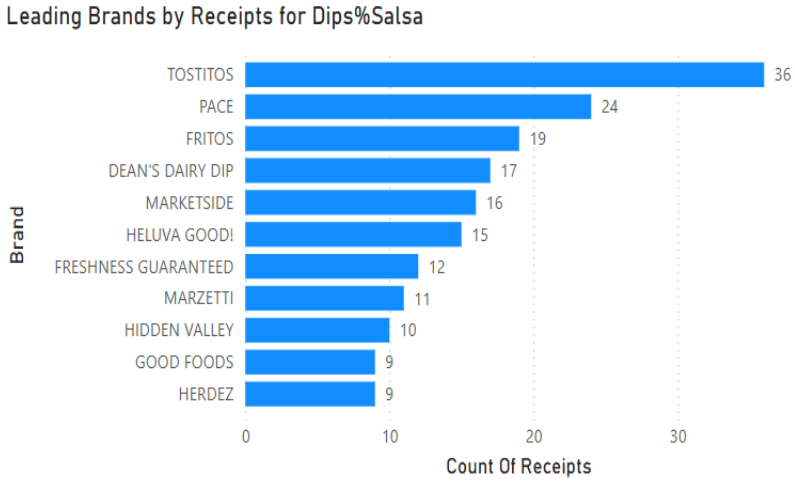
(Date.Year(DateTime.LocalNow()) - Date.Year([CREATED_DATE])) * 12 +
(Date.Month(DateTime.LocalNow()) - Date.Month([CREATED_DATE]))

Business Questions in Power BI:

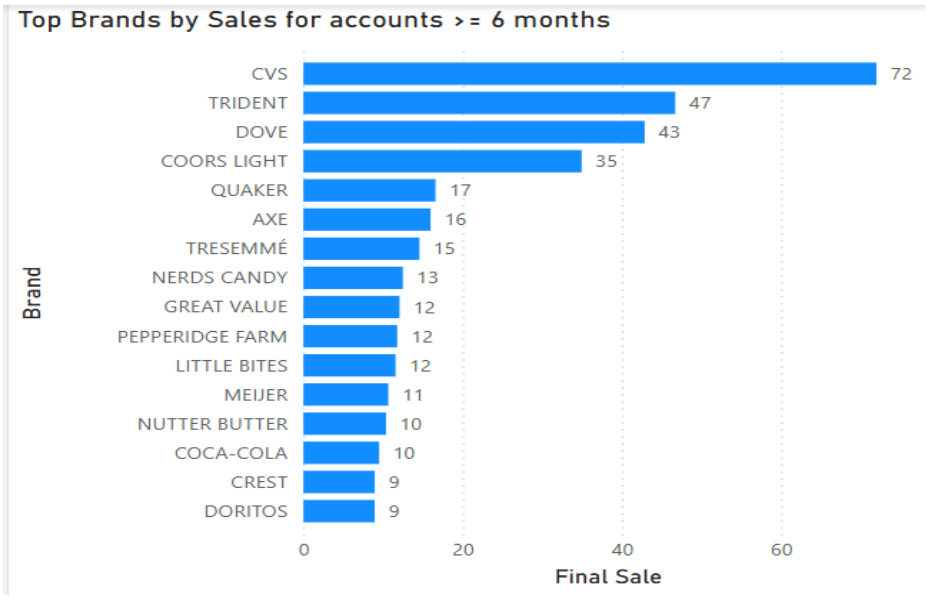
3)At what percent has Fetch grown year over year? (Open-ended question)



2) Which is the leading brand in the Dips & Salsa category?(Open ended Question)



2) What are the top 5 brands by sales among users that have had their account for at least six months? (Close ended Question)



1) What are the top 5 brands by receipts scanned among users 21 and over?(Close ended Question)