Close Ended Questions:

1. **What are the top 5 brands by receipts scanned among users 21 and over?**
   Ans. Here I'm assuming by receipts scanned means total number of receipts scanned per brand. I'm calculating the brand and their respective receipts scanned for users who are or above 21 years old. Also added a where condition to discard null brands.

```
WITH users21AndOlder as
        (select id
                from users
                where (CURRENT_DATE - DATE(birth_date))/ 365.25 >= 21)
select
         p.brand
        ,count(t.receipt_id) as transactions_count
from transactions t
        join users21AndOlder u on t.user_id = u.id
        join products p on p.barcode = t.barcode
where p.brand is not null
group by p.brand
order by count(t.receipt_id) desc
limit 5;
```

Query Editor

```
1  WITH users21AndOlder as
2  (select id from users where (CURRENT_DATE - DATE(birth_date))/ 365.25 >= 21)
3  select
4       p.brand
5      ,count(t.receipt_id) as transactions_count
6  from transactions t
7      join users21AndOlder u on t.user_id = u.id
8      join products p on p.barcode = t.barcode
9  where p.brand is not null
10 group by p.brand
11 order by count(t.receipt_id) desc
12 limit 5;
```

Explain | Notifications | Messages | Data Output | Query History

	brand character varying (100)	transactions_count bigint
1	NERDS CANDY	6
2	DOVE	6
3	MEIJER	4
4	GREAT VALUE	4
5	COCA-COLA	4

2. **What are the top 5 brands by sales among users that have had their account for at least 6 months?**

Ans.

```
WITH usersActiveFor6orMoreMonths as
        (select id
        from users
        where (CURRENT_DATE - DATE(created_date))>180 )
select p.brand
          ,sum(t.sale) as total_sales
from transactions t
        join usersActiveFor6orMoreMonths u on u.id = t.user_id
        join products p on p.barcode = t.barcode
where p.brand is not null
group by p.brand
order by sum(t.sale) desc
limit 5;
```

Query Editor

```
14
15
16  WITH usersActiveFor6orMoreMonths as
17  (select id from users where (CURRENT_DATE-DATE(created_date))>180 )
18  select p.brand
19         ,sum(t.sale) as total_sales
20  from transactions t
21      join usersActiveFor6orMoreMonths u on u.id = t.user_id
22      join products p on p.barcode = t.barcode
23  where p.brand is not null
24  group by p.brand
25  order by sum(t.sale) desc
26  limit 5;
27
```

Explain | Notifications | Messages | Data Output | Query History

	brand character varying (100)	total_sales numeric
1	CVS	72.0000
2	TRIDENT	46.7200
3	DOVE	42.8800
4	COORS LIGHT	34.9600
5	QUAKER	16.6000

Open Ended Question:

2. Leading brand in the Dips and Salsa category?

Ans. I'm defining a leading brand as the brand which has a higher number of transactions and higher sales as well. I tried to find if there is a brand with less number of transactions but higher sales and there was not. In both cases there is one clear winner that is 'Tostitos'.

```
select distinct p.brand
                ,count(t.receipt_id) as transaction_count
                ,sum(t.sale) as total_sales
from products p
join transactions t on p.barcode = t.barcode
where p.category_2 = 'Dips & Salsa'
group by p.brand
order by sum(t.sale) desc;
```

```
30
31  select distinct p.brand
32          ,count(t.receipt_id) as transaction_count
33          ,sum(t.sale) as total_sales
34  from products p
35  join transactions t on p.barcode = t.barcode
36  where p.category_2 = 'Dips & Salsa'
37  group by p.brand
38  order by sum(t.sale) desc;
39
40
41
```

Explain | Notifications | Messages | Data Output | Query History

	brand character varying (100)	transaction_count bigint	total_sales numeric
1	TOSTITOS	72	260.9900
2	[null]	42	154.3700
3	GOOD FOODS	18	118.8900
4	PACE	48	118.5800
5	MARKETSIDE	32	103.2900
6	FRITOS	38	91.7300
7	HELUVA GOOD!	30	85.4600
8	FRESHNESS GUARANTEED	24	73.3700

3. Year over year growth for Fetch.

Ans. There is very limited data available to find the year over year growth for Fetch. For this question, I'll be only looking at the number of customers Fetch has gained year over year. I'm using the users table and assuming all these users are active users, since there is no current way to figure out the exact active users since the transactions table also only has 3 months of data.

First I'm finding the new years in each year and also the total users using the cumulative sum. The formula I'm using to find YoY growth is
(this_year_new_users_count / Last_year_total_users_count)* 100.

Query:

```
with new_users_yoy as(
		select
			date_part('year', created_date) as joining_year
			,count(id) as new_users_added
		from users
		group by date_part('year', created_date)
		order by date_part('year', created_date)
	),
	total_users_yoy as(
		select joining_year
			,new_users_added
			,sum(new_users_added) over (order by joining_year) as total_users
			from new_users_yoy
	)
select joining_year
		,new_users_added
		,total_users
		,ROUND((new_users_added/lag(total_users) over (order by joining_year))*100,
2) as "yoy_growth_%"
		from total_users_yoy
```

Output:

Query Editor

```
with new_users_yoy as(
        select
            date_part('year', created_date) as joining_year
            ,count(id) as new_users_added
        from users
        group by date_part('year', created_date)
        order by date_part('year', created_date)
    ),
    total_users_yoy as(
        select joining_year
            ,new_users_added
            ,sum(new_users_added) over (order by joining_year) as total_users
            from new_users_yoy
    )
select joining_year
        ,new_users_added
        ,total_users
        ,ROUND((new_users_added/lag(total_users) over (order by joining_year))*100, 2) as "yoy_growth_%"
        from total_users_yoy
```

Explain Notifications Messages Data Output Query History

	joining_year double precision	new_users_added bigint	total_users numeric	yoy_growth_% numeric
1	2014	30	30	[null]
2	2015	51	81	170.00
3	2016	70	151	86.42
4	2017	644	795	426.49
5	2018	2168	2963	272.70
6	2019	7093	10056	239.39
7	2020	16883	26939	167.89
8	2021	19159	46098	71.12
9	2022	26807	72905	58.15
10	2023	15464	88369	21.21
11	2024	11631	100000	13.16

The next step is simple to calculate the average YoY growth, the output as follows. **Based on assumptions I made and looking at only the new customer base acquired each year, it is safe to say Fetch grew 152.65% YoY over the last 10 years.**

Query Editor

```
with new_users_yoy as(
        select
            date_part('year', created_date) as joining_year
            ,count(id) as new_users_added
        from users
        group by date_part('year', created_date)
        order by date_part('year', created_date)
    ),
    total_users_yoy as(
        select joining_year
            ,new_users_added
            ,sum(new_users_added) over (order by joining_year) as total_users
            from new_users_yoy
    ),
    yoy_growth as
    (select joining_year
        ,new_users_added
        ,total_users
        ,ROUND((new_users_added/lag(total_users) over (order by joining_year))*100, 2) as "yoy_growth_%"
        from total_users_yoy)
    select round(avg("yoy_growth_%"), 2) as avg_YoY_growth_percent
    from yoy_growth
    where "yoy_growth_%" is not null


```

Explain Notifications Messages Data Output Query History

	avg_yoy_growth_percent numeric
1	152.65