Closed-ended questions:

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

```
with user_age as (
        select ID, datediff(year, BIRTH_DATE, GETDATE()) as age from USER_TAKEHOME),
result as (
        select p.BRAND, t.RECEIPT_ID
        from TRANSACTION_TAKEHOME t
        join user_age u on t.USER_ID = u.ID
        join PRODUCT_TAKEHOME p on t.BARCODE = p.BARCODE
        where u.age >= 21 and p.BRAND is not null)
select TOP 5 BRAND, count(RECEIPT_ID) as total_receipts
from result
group by BRAND
order by total_receipts desc;
```

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

```
with user_account_age as (
        select ID, datediff(month, CREATED_DATE, GETDATE()) as account_age
        from USER_TAKEHOME),
result as (
        select p.BRAND, t.FINAL_SALE
        from TRANSACTION_TAKEHOME t
        join user_account_age u on t.USER_ID = u.ID
        join PRODUCT_TAKEHOME p on t.BARCODE = p.BARCODE
        where u.account_age >= 6 and p.BRAND is not null)
select TOP 5 BRAND, sum(FINAL_SALE) as total_sales
from result
group by BRAND
order by total_sales desc;
```

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

```
With generation_data as (
        select
                case
                        when year(u.BIRTH_DATE) <1965 then 'Baby Boomer  '
                        when year(u.BIRTH_DATE) between 1965 and 1980 then 'GenerationX'
                        when year(u.BIRTH_DATE) between 1981 and 1996 then 'Millenials'
                        when year(u.BIRTH_DATE) between 1997 and 2012 then 'Generation Z'
                        else 'Generation Alpha'
                end as generation,
                t.FINAL_SALE
        from TRANSACTION_TAKEHOME t
        join USER_TAKEHOME u on t.USER_ID = u.ID
        join PRODUCT_TAKEHOME p on t.BARCODE = p.BARCODE
        where p.CATEGORY_1 = 'Health & Wellness')
select
        generation,
        (sum(FINAL_SALE)*100 / (select sum(FINAL_SALE) from generation_data)) as
sales_percentage
from  generation_data
group by generation
order by sales_percentage desc;
```

Open-ended questions:

- **Who are Fetch's power users?**

```
With generation_data as (
        select
                case
                        when year(u.BIRTH_DATE) <1965 then 'Baby Boomer  '
                        when year(u.BIRTH_DATE) between 1965 and 1980 then 'GenerationX'
                        when year(u.BIRTH_DATE) between 1981 and 1996 then 'Millenials'
                        when year(u.BIRTH_DATE) between 1997 and 2012 then 'Generation Z'
                        else 'Generation Alpha'
                end as generation,
                t.RECEIPT_ID
```

```
        from TRANSACTION_TAKEHOME t
        join USER_TAKEHOME u on t.USER_ID = u.ID
        where u.BIRTH_DATE is not null)
select
        TOP 1 generation,
        (count(RECEIPT_ID)*100 / (select count(RECEIPT_ID) from generation_data)) as
usage_percent
from  generation_data
group by generation
order by usage_percent desc;
```

This above query gives us the which generation people are power user's for Fetch. And below are assumptions:

1) Power user is person with more transactions (more receipts scanned)
2) Datasets doesn't have null or duplicate rows.
3) The years of generation is as assumed above.
4) ID is primary key in USER_TAKEHOME dataset.

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

```
with category_data as (
        select distinct BARCODE, BRAND
        from PRODUCTS_TAKEHOME
        where CATEGORY_2  = 'Dips & Salsa'
         and BARCODE is not null
         and BRAND is not null),
c_sales as (
        select c.BRAND, t.FINAL_SALE
        from TRANSACTION_TAKEHOME t
        join category_data c on t.BARCODE = c.BARCODE)
select TOP 1 BRAND, sum(FINAL_SALE) as total_sales
from c_sales
group by BRAND
order by sum(FINAL_SALE) desc;
```

The assumptions in the above case are:

1) Datasets doesn't have duplicates rows
2) BARCODE and BRAND columns doesn't have null values in PRODUCT_TAKEHOME dataset and BARCODE column is primary key.
3) BARCODE, FINAL_SALE columns are not null in TRANSACTION_TAKEHOME dataset