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

```python
# Q. What are the top 5 brands by receipts scanned among users 21 and over?


# Assumptions:
# 1. Not considering records where the brand name is unknown or 'nan'
# Reasoning: We don't know if the brands which are marked as 'nan' are same or
different brands
# Grouping it as 1 brand and showing it on top 5 does not make sense

query1 = """
WITH valid_users as (SELECT
    receipt_id,
    strftime('%Y', 'now') - strftime('%Y', birth_date) - (strftime('%m-%d', 'now') <
strftime('%m-%d', birth_date)) AS Age,
    brand
FROM users u
INNER JOIN transactions t on t.user_id = u.id
INNER JOIN products p on p.barcode = t.barcode)

SELECT brand as "Brand name", count(receipt_id) as "Number of receipts"
FROM valid_users
WHERE age >= 21 AND brand != 'nan'
GROUP BY brand
ORDER BY count(receipt_id) DESC
LIMIT 5
;
"""


result1 = pd.read_sql(query1, conn)
print(result1)
```

OUTPUT:

	Brand name	Number of receipts
0	NERDS CANDY	6
1	DOVE	6
2	TRIDENT	4
3	SOUR PATCH KIDS	4
4	MEIJER	4

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

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

# Assumptions:
# Not considering brands with unknown value or 'nan' value
# Included the rows where final_quantity is 0, because in many rows even though
quantity is 0 the sale value is there


query2 = """

WITH eligible_users as (
  SELECT
      id,
      (strftime('%Y', 'now') - strftime('%Y', created_date)) * 12 +
      (strftime('%m', 'now') - strftime('%m', created_date)) AS
account_lifetime_months
  FROM users
),

sales as (
  SELECT
      id,
      final_sale,
      brand
  FROM eligible_users u
  INNER JOIN transactions t on u.id = t.user_id
  INNER JOIN products p on p.barcode = t.barcode
  WHERE account_lifetime_months >= 6 and brand != 'nan'
)

SELECT brand, sum(final_sale) as "Total sales"
FROM sales
GROUP BY brand
ORDER BY sum(final_sale) DESC
LIMIT 5
;

"""
```

```
result2 = pd.read_sql(query2, conn)
print(result2)
```

OUTPUT:

```
        brand  Total sales
0         CVS        72.00
1     TRIDENT        46.72
2        DOVE        42.88
3  COORS LIGHT       34.96
4      QUAKER        16.60
```

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

```
# Q. What is the percentage of sales in the Health & Wellness category by generation?
#
https://www.parents.com/parenting/better-parenting/style/generation-names-and-years-a-
cheat-sheet-for-parents/
# Taking reference for generation from above link

query3 = """
WITH valid_transactions as (
    SELECT
        *
    FROM products p
    INNER JOIN transactions t on t.barcode = p.barcode
    INNER JOIN users u on u.id = t.user_id
    WHERE category_1 = 'Health & Wellness'
),

gen_users as (
    SELECT
        id,
        CASE
            when strftime('%Y', birth_date) BETWEEN '1900' AND '1927' then 'Greatest
Generation'
            when strftime('%Y', birth_date) BETWEEN '1928' AND '1945' then 'Silent
Generation'
            when strftime('%Y', birth_date) BETWEEN '1946' AND '1964' then 'Baby Boom
Generation'
```

```
            when strftime('%Y', birth_date) BETWEEN '1965' AND '1980' then 'Generation
X'
            when strftime('%Y', birth_date) BETWEEN '1981' AND '1996' then 'Millennial'
            when strftime('%Y', birth_date) BETWEEN '1997' AND '2010' then 'Generation
Z'
            when strftime('%Y', birth_date) BETWEEN '2011' AND '2024' then 'Generation
Alpha'
            END Generation
    FROM users
),

total_sales as (
    SELECT
        SUM(final_sale) as total_health_wellness_sales
    FROM valid_transactions
)

SELECT
    generation,
    ROUND((SUM(final_sale) / (select total_health_wellness_sales from total_sales)) *
100.0, 2) as "Health & Wellness Sales %"
FROM valid_transactions v
INNER JOIN gen_users g on v.user_id = g.id
GROUP BY generation
ORDER BY "Health & Wellness Sales %" DESC
;
"""
result3 = pd.read_sql(query3, conn)
print(result3)
```

OUTPUT:

```
                 Generation  Health & Wellness Sales %
0  Baby Boom Generation                          46.94
1             Millennial                          31.18
2            Generation X                          21.88
```