Research Request Analyst
SQL Questionnaire

1. Suppose a retail chain has the two tables shown below, a table of store locations and a table of annual revenue for each store. Write a query to produce a new table which shows each store location on the same row as the annual revenue for that store.

[dbo].[STORE_LOCATION_TABLE]

STORE_ID	STORE_LOCATION	UPDATE_DATE
1	123 Cherry Ave.	2024-08-16
2	456 Peach St.	1997-02-07
2	456 East Peach St.	2024-08-16
3	789 Mango Rd.	2024-08-16
4	246 Plum Ln.	2024-08-16

[dbo].[STORE_REVENUE_TABLE]

STORE_ID	ANNUAL_REVENUE
1	$120,000
2	$340,000
3	$567,000
4	$890,000

```
- - Getting store location and its revenue

SELECT
   loc.STORE_ID,
   loc.STORE_LOCATION,
   rev.ANNUAL_REVENUE

FROM
   [dbo].[STORE_LOCATION_TABLE] loc

JOIN
```

```
   [dbo].[STORE_REVENUE_TABLE] rev

ON
   loc.STORE_ID = rev.STORE_ID;
```

2. Explain how you would calculate the sum of the annual revenue for all the stores using the results table from your query in Question 3. Will your sum be an accurate representation of the total revenue of the stores?

To calculate the total revenue for all stores, I would use SUM() Function:

```
-- Calculating the total revenue across all stores

SELECT SUM(ANNUAL_REVENUE) AS TOTAL_REVENUE
FROM (

   -- Ensuring unique store revenue records

   SELECT DISTINCT loc.STORE_ID, rev.ANNUAL_REVENUE
   FROM [dbo].[STORE_LOCATION_TABLE] loc
   JOIN [dbo].[STORE_REVENUE_TABLE] rev
   ON loc.STORE_ID = rev.STORE_ID
) AS RevenueData;
```

3. Confabulous Confections, Inc. produces premium candies and ships them around the country. Lemon drops, candy canes, mocha marshmallows, strawberry truffles, bourbon bonbons, etc. However, the shipping service will not deliver the bourbon bonbons to Utah (UT) due to regulations. Suppose the orders are stored in a table like this:

[dbo].[ORDER_TABLE]

ORDER_ID	ITEM_DESC	DELIVERY_CITY	DELIVERY_STATE
1	lemon drops	Chicago	IL
2	bourbon bonbons	Denver	CO
3	candy canes	Los Angeles	CA

4	bourbon bonbons	Salt Lake City	UT
5	mocha marshmallows	Atlanta	GA
6	strawberry truffles	Houston	TX
7	candy canes	Provo	UT
etc.	*etc.*	*etc.*	*etc.*

Below is an incomplete query to pull only the orders which the shipping service will deliver. Complete the WHERE clause appropriately.

```
-- Fetching all orders except bourbon bonbons shipping to Utah

SELECT *
FROM [dbo].[ORDER_TABLE]
WHERE ITEM_DESC != 'bourbon bonbons'
  OR DELIVERY_STATE != 'UT';
```

4. Suppose we have a large table listing every household in the United States in the following format:

[dbo].[US_HOUSEHOLDS_TABLE]

HOUSEHOLD_ID	FAMILY_NAME	MEMBER_NAME
1	Anderson	Robert
1	Anderson	Mary
1	Anderson	Cody
1	Anderson	Brittany
2	Carver	William
2	Carver	Suzanne
3	Washington	Fred
3	Washington	Ethel
3	Washington	Tony
4	Smith	Darlene
4	Smith	Amy

4	Smith	Richard
etc.	*etc.*	*etc.*

Write a query to aggregate the table above to produce a frequency table of family size, similar to the following:

FAMILY_SIZE	FREQUENCY
1	23,903,124
2	57,116,548
3	38,617,295
4	12,397,106
5	3,108,284
6 or more	1,393,571

```
-- Counting the number of households based on family size

SELECT
    family.FAMILY_SIZE,
    COUNT(*) AS FREQUENCY
FROM (

    -- Getting family size for each household

    SELECT HOUSEHOLD_ID, COUNT(*) AS FAMILY_SIZE
    FROM [dbo].[US_HOUSEHOLDS_TABLE]
    GROUP BY HOUSEHOLD_ID
) AS family
GROUP BY family.FAMILY_SIZE
ORDER BY family.FAMILY_SIZE;
```

To group families of 6 or more together:

```
-- Categorizing households into specific family size groups

SELECT
    CASE
```

```
        WHEN FAMILY_SIZE >= 6 THEN '6 or more'
        ELSE CAST(FAMILY_SIZE AS VARCHAR)
    END AS FAMILY_SIZE_GROUP,
    COUNT(*) AS FREQUENCY
FROM (

    -- Determining the number of members per household

    SELECT HOUSEHOLD_ID, COUNT(*) AS FAMILY_SIZE
    FROM [dbo].[US_HOUSEHOLDS_TABLE]
    GROUP BY HOUSEHOLD_ID
) AS family
GROUP BY
    CASE
        WHEN FAMILY_SIZE >= 6 THEN '6 or more'
        ELSE CAST(FAMILY_SIZE AS VARCHAR)
    END
ORDER BY FAMILY_SIZE_GROUP;
```