Scenario:

Data Scientist at USDA (United States Department of Agriculture)

Context:

You are a Data Scientist working at the USDA. Your department has been tracking the production of various agricultural commodities across different states.

Your datasets include:

`milk_production`, `cheese_production`, `coffee_production`, `honey_production`, `yogurt_production`, and a `state_lookup` table.

The data spans multiple years and states, with varying levels of production for each commodity.

Your manager has requested that you generate insights from this data to aid in future planning and decision-making. You'll need to use SQL queries to answer the questions that come up in meetings, reports, or strategic discussions.

Objectives:

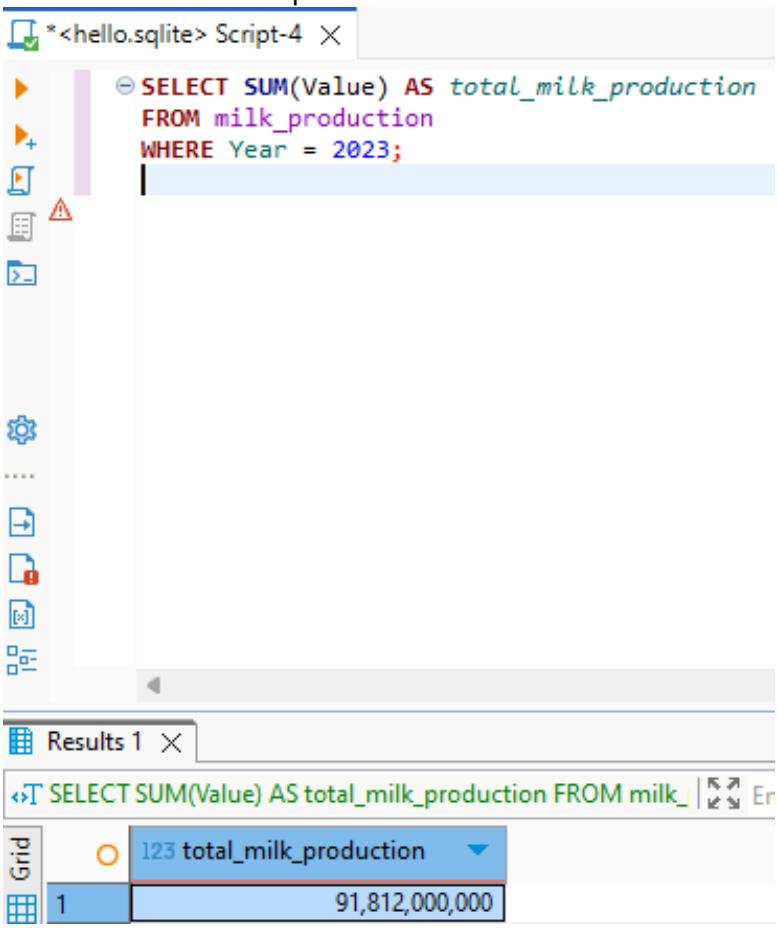
- Assess state-by-state production for each commodity.
- Identify trends or anomalies.
- Offer data-backed suggestions for areas that may need more attention.

NOTE: All answer entries are numeric and only numbers and periods. The autograder does not accept commas for the final project.

1. Question 1

Can you find out the total milk production for 2023? My manager wants this information for the yearly report.

What is the total milk production for 2023?



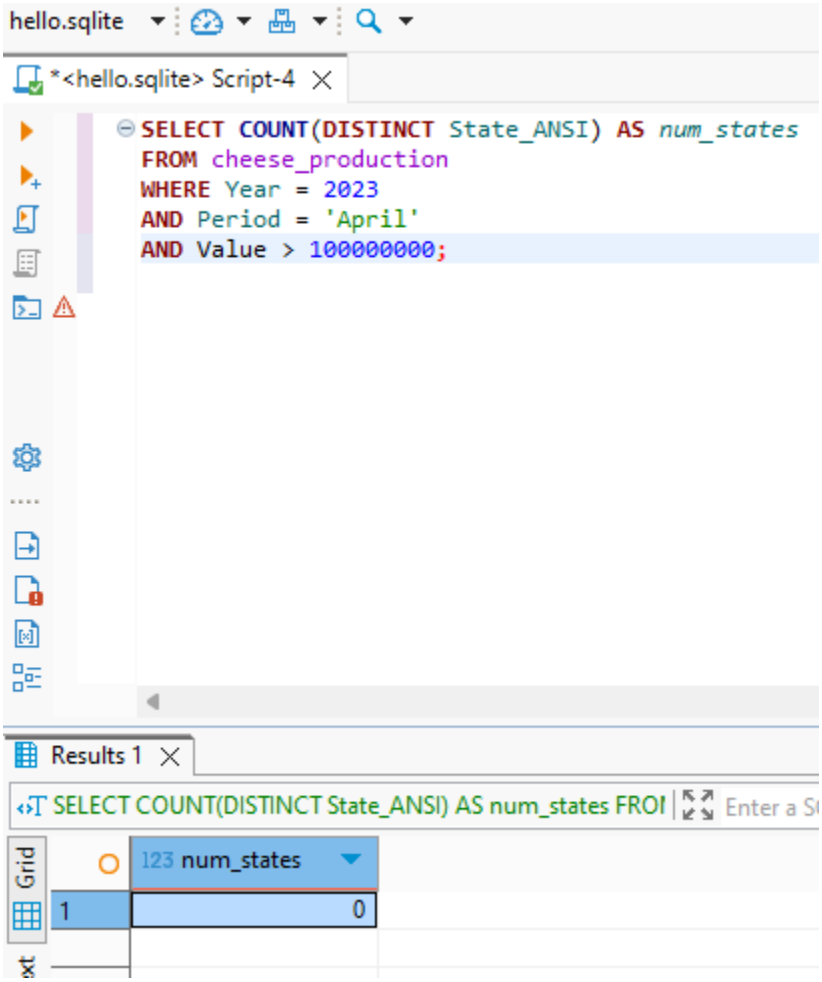
Analyst Explanation:

- SUM(Value): Calculates the **total milk production**.
- WHERE Year = 2023: Filters the records to only include production data from **2023**.

2. Question 2

Which states had cheese production greater than 100 million in April 2023? The Cheese Department wants to focus their marketing efforts there.

How many states are there?



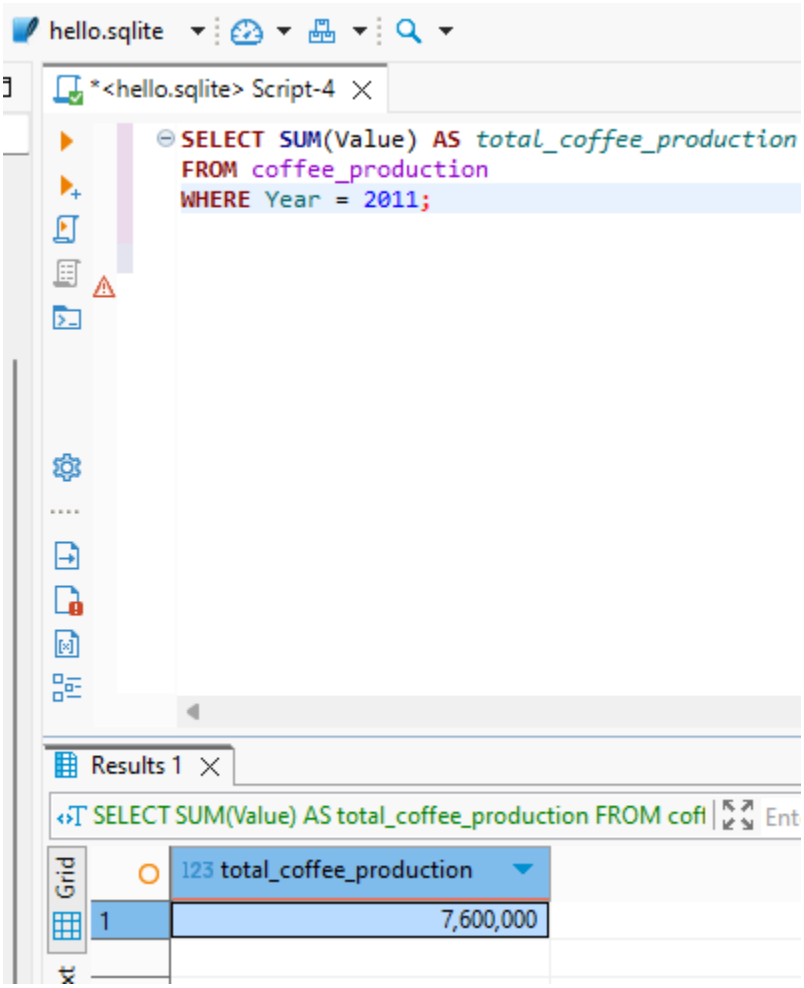
Analyst Explanation:

- `COUNT(DISTINCT State_ANSI)`: Counts the **number of unique states** meeting the criteria.
- `WHERE Year = 2023 AND Period = 'April'`: Filters data for **April 2023**.
- `AND Value > 100000000`: Ensures only states where cheese production **exceeded 100 million** are included.

--

3. Question 3
   Your manager wants to know how coffee production has changed over the years.

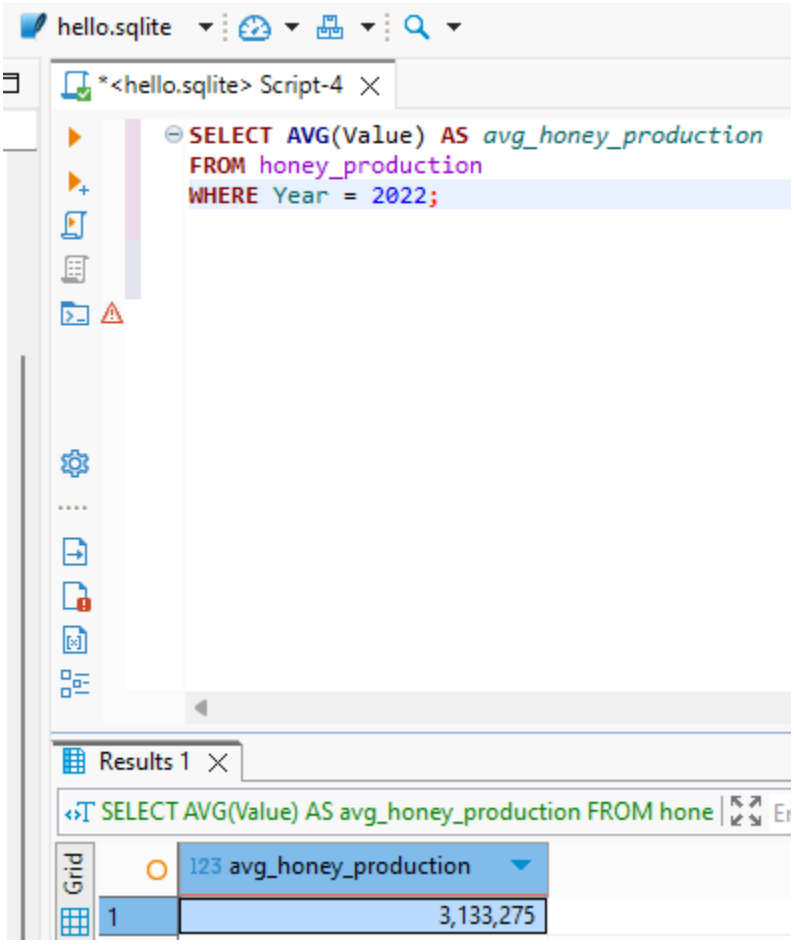
What is the total value of coffee production for 2011?



-

4. Question 4

There's a meeting with the Honey Council next week. Find the average honey production for 2022 so you're prepared.



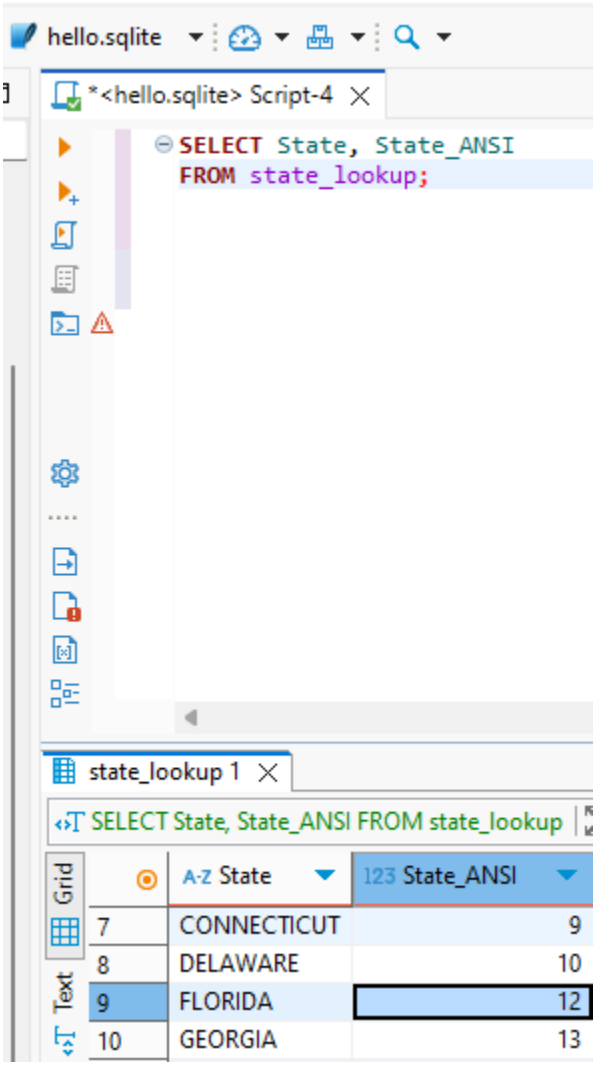
Analyst Explanation:

- `AVG(Value)`: Calculates the **average honey production**.
- `WHERE Year = 2022`: Filters records to only include **2022** data.

Question 5

The State Relations team wants a list of all states names with their corresponding ANSI codes. Can you generate that list?

What is the State_ANSI code for Florida?



Analyst Explanation:

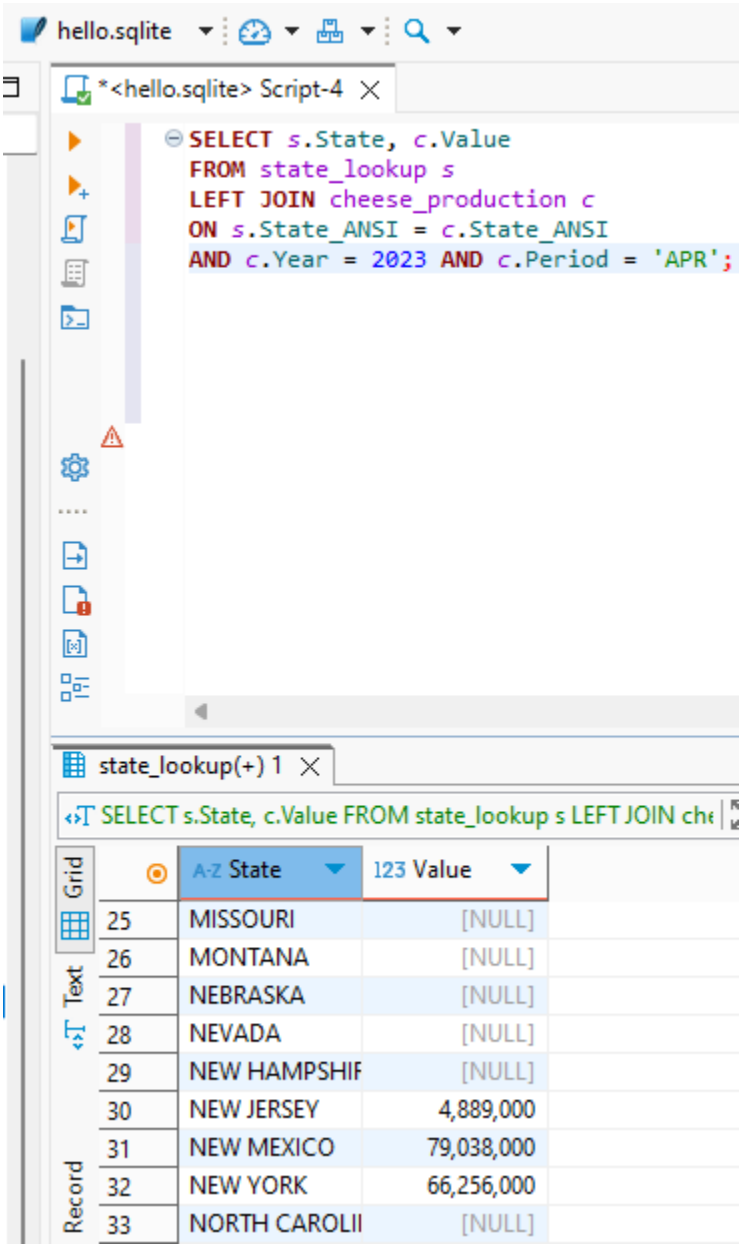
- **SELECT State, State_ANSI** → Selects the columns containing the state name and ANSI code.

- **FROM state_lookup** → Ensures the data is pulled from the correct reference table that stores state information.

6. Question 6

For a cross-commodity report, can you list all states with their cheese production values, even if they didn't produce any cheese in April of 2023?

What is the total for NEW JERSEY?



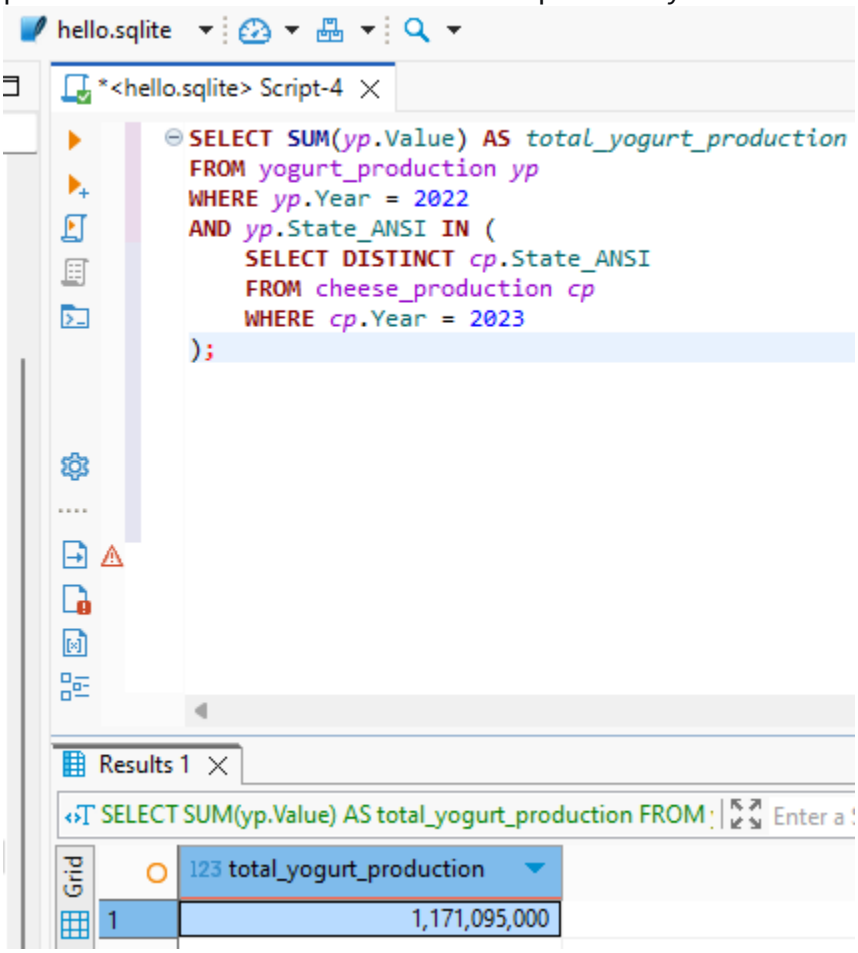
Analyst Explanation:

- `LEFT JOIN state_lookup s ON cheese_production c` → Ensures all states are included, even if they **didn't produce cheese** in April 2023.
- `AND c.Year = 2023 AND c.Period = 'APR'` → Filters only **2023 April cheese production** records.

- `s.State, c.Value` → Displays **state names** alongside **cheese production values**.

7. Question 7

Can you find the total yogurt production for states in the year 2022 which also have cheese production data from 2023? This will help the Dairy Division in their planning.



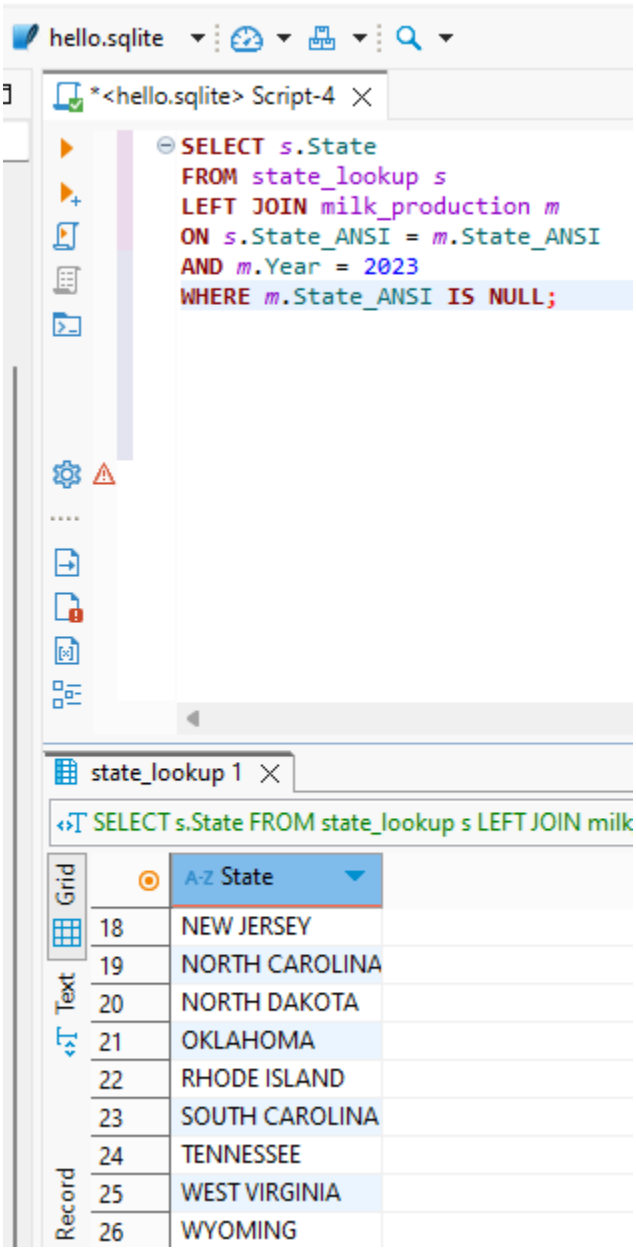
Analyst Explanation:

- `SUM(yp.Value)`: Calculates the **total yogurt production** for qualifying states.
- `WHERE yp.Year = 2022`: Filters yogurt production records to only include **2022**.
- **Subquery (`IN (...)`)**: Retrieves **distinct** states (`State_ANSI`) that have **cheese production data from 2023**, ensuring only those states are included in the calculation.

8. Question 8

List all states from state_lookup that are missing from milk production in 2023.

How many states are there?



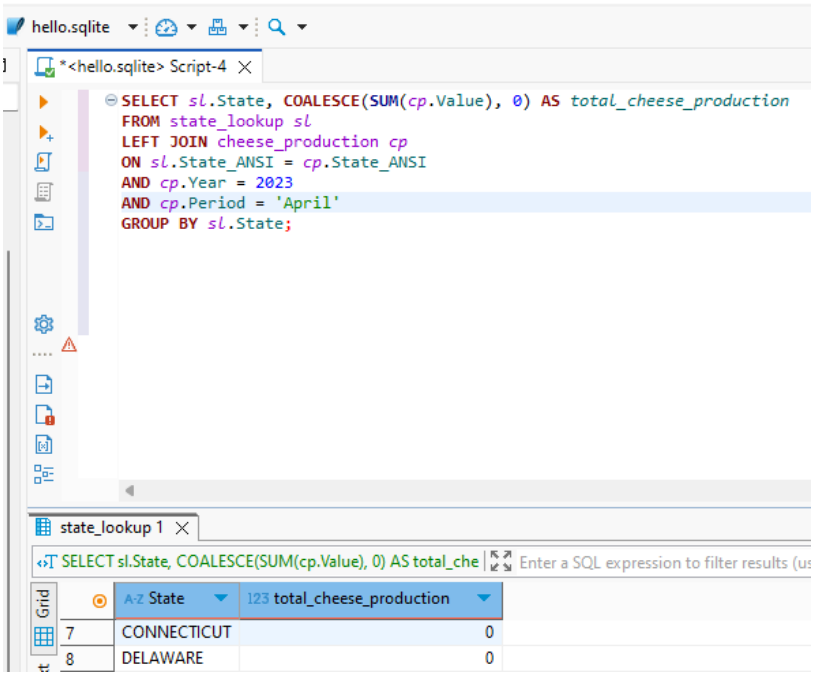
Question 9

List all states with their cheese production values, including states that didn't produce any cheese in April 2023.

Did Delaware produce any cheese in April 2023?

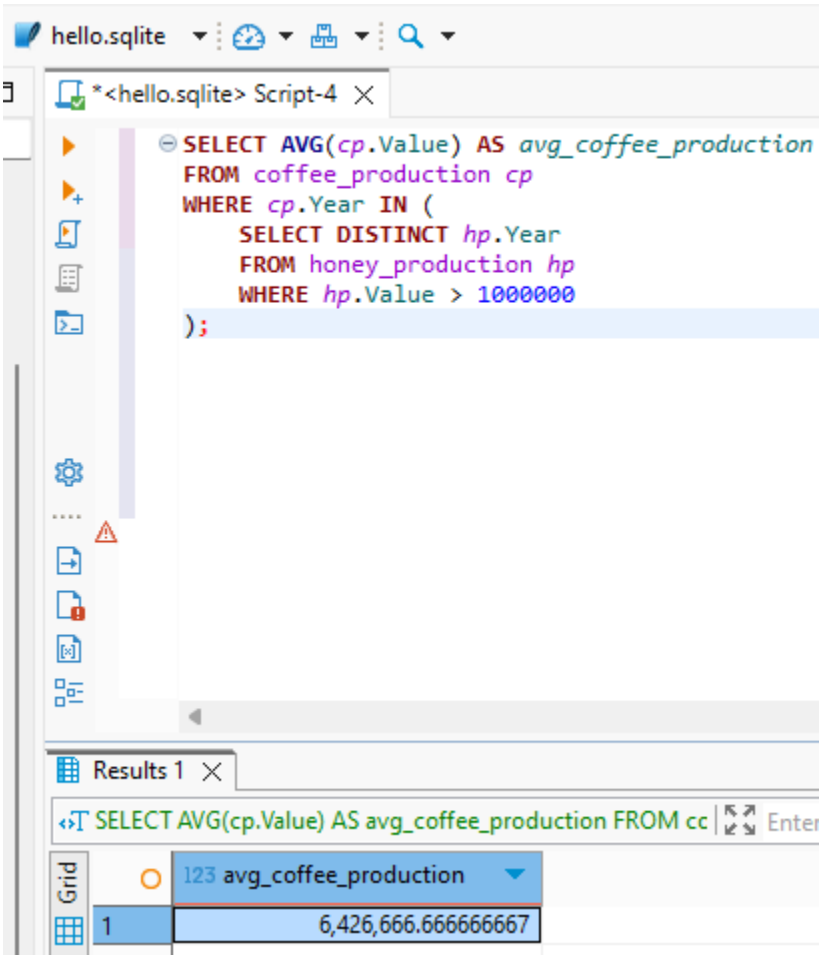
Analyst Explanation:

- `LEFT JOIN` ensures all states from `state_lookup` are included, even if they **didn't produce cheese** in April 2023.
- `COALESCE(SUM(cp.Value), 0)` replaces `NULL` values with `0` for states that didn't report cheese production.
- `GROUP BY sl.State` ensures each state appears only once in the results.



Question 10

Find the average coffee production for all years where the honey production exceeded 1 million.



Analyst Explanation:

- `AVG(cp.Value)`: Computes the **average coffee production**.
- `WHERE cp.Year IN (...)`: Filters coffee production to only include years where honey production exceeded **1 million**.
- **Subquery (`IN (...)`)**: Retrieves **distinct years (`Year`)** from `honey_production` where `Value > 1000000`.