Scenario:

SQL in data science at USDA (United States Department of Agriculture)

Context:

I am working at the USDA. The department has been tracking the production of various agricultural commodities across different states.

The 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.

Our manager has requested that we generate insights from this data to aid in future planning and decision-making. I 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.

1.

Query 1

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

What is the total milk production for 2023? 91812000000

```
SELECT SUM(value) AS total_milk_production
FROM Milk_Production
WHERE year = 2023
```

2.

Query 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? 2

```
SELECT s.state, c.value
```

FROM Cheese_Production c

JOIN State_Lookup s ON c.state_ansi = s.state_ansi

WHERE c.year = 2023 AND c.period= 'apr' AND c.value > 100000000

3.

Query 3

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

What is the total value of coffee production for 2011? 7600000

SELECT SUM(value) AS total_coffee_production

FROM Coffee_Production

WHERE year = 2011

4.

Query 4

There's a meeting with the Honey Council next week. Find the average honey production for 2022 so you're prepared. 3133275

SELECT AVG(value) AS average_honey_production

FROM Honey_Production

WHERE year = 2022

5.

Query 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? 12

SELECT state_ansi

FROM State_Lookup

WHERE state = 'Florida'

6.

Query 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? 4889000

```
SELECT COALESCE(SUM(c.value), 0) AS total_cheese_production_new_jersey
FROM Cheese_Production c
JOIN State_Lookup s ON c.state_ansi = s.state_ansi
WHERE s.state = 'New Jersey'
  AND c.year = 2023
  AND c.period = 'apr'
```

7.

Query 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. 1171094999

```
SELECT sum(y.Value) AS total_yogurt_production
FROM Yogurt_Production y
WHERE y.year = 2022
  AND y.state_ansi IN (
    SELECT DISTINCT c.State_ANSI
    FROM Cheese_Production c
    WHERE c.year = 2023
  )
```

8.

Query 8

List all states from state_lookup that are missing from milk_production in 2023.

How many states are there? 26

```
 SELECT s.state
FROM state_lookup s
LEFT JOIN milk_production m ON s.state_ansi = m.state_ansi AND m.year = 2023
WHERE m.state_ansi IS NULL
```

9.

Query 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? No

```
SELECT sl.state_ansi , sl.state, cp.value
 from state_lookup sl
 left join cheese_production cp on cp.State_ANSI = sl.State_ANSI and year = 2023
 where Period = 'APR' and cp.value is null
```

10.

Query 10

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

```
select c.year, avg (c.value) avg_coffe_value , h.Value
 from coffee_production c
 left join honey_production h
 on c.State_ANSI = h.State_ANSI
 group by c.Year, h.Value
 having h.Value >1000000
```