Data Analytics - Career Accelerator

SQL
Gym Session

Masterschool

Data Analytics - Career Accelerator

SQL Gym Session

While we're waiting to begin,
connect to the database via Beekeeper
using the connection string in the Zoom
chat.

SQL Gym Session

What we'll do today

- Work individually on a number of exercises in SQL, progressing from easy to difficult
- Review solutions as a group after each question

FAOSTAT Dataset

- Using Beekeeper, connect to the database using the connection URL (links in the Zoom chat)
- Crops and livestock dataset published by the Food & Agriculture Organization (FAO) of the United Nations (UN)
- Take a minute to get familiar with the data

Agricultural Metrics

Area harvested: the total area harvested for crops

Producing Animals/Slaughtered/Stock: the number of animals for livestock/animal products

Production: the total weight of the crops/livestock/animal product produced

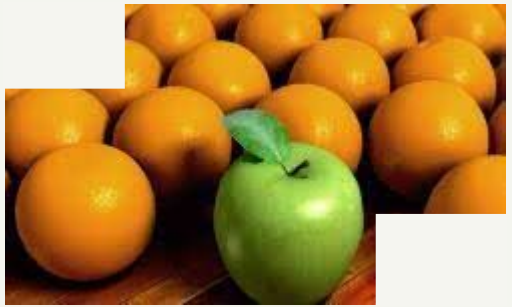
Yield: the weight produced per unit of area harvested or animal

"Easy" Questions

Question 1

Question 1

How many different elements, countries, and years are there?

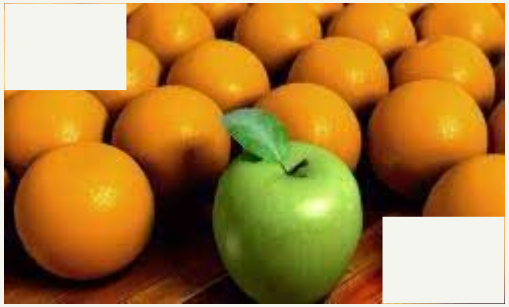


Question 1 - Solution

Question 1 - Solution

How many different elements, countries, and years are there?

```
SELECT
     COUNT (DISTINCT element) AS num_of_activities,
     COUNT (DISTINCT area) AS num_of_countries,
     COUNT (DISTINCT year)AS num_of_year
FROM crop_livestock_stats;
```



Question 2

Question 2

What's the total number of animals slaughtered?

Question 2 - Solution

Question 2

What's the total number of animals slaughtered?

→ 244,427

Possible Solution:

```
SELECT COUNT(*) AS num_of_slaughters
FROM crop_livestock_stats
WHERE element = 'Producing Animals/Slaughtered';
```

"Medium" Questions

Question 3

Question 3

For Canada, the US, and Mexico, beginning after the year 2010, return the average yield by item where the average is > 380000 .

Question 3 - Solution

Question 3 - Solution

For Canada, the US, and Mexico beginning after the year 2010, return the average yield by item where the average is > 380000 .

```
SELECT year, area AS country, item,
     AVG(value) AS avg_yield, element
FROM crop_livestock_stats
WHERE area IN ('Canada', 'United States of America',
     'Mexico') AND element = 'Yield' AND year > 2010
GROUP BY area, item, element, year
HAVING AVG(value) > 380000
ORDER BY avg_yield DESC;
```

Question 4

Question 4

What is the yearly quantity of each product, measured in tonnes, for each country ranked from largest to smallest quantities? HINT: USE JOIN

Question 4 - Solution

Question 4 - Solution

What is the yearly quantity of each product in tonnes for each country ranked from largest to smallest quantities? HINT: USE JOIN

Possible Solution:

```
SELECT s.year, s.area AS country,
     s.item AS produce,
     SUM(s.value) AS total_weight, s.unit,
     c.country
FROM crop_livestock_stats AS s
LEFT JOIN country_groups AS c
ON s.area_code = c.country_group_code
WHERE s.unit = 'tonnes'
     and c.country_group_code is null
GROUP BY s.year, s.area, s.item, s.unit, c.country
ORDER BY total_weight DESC;
```

3-minute break

"Hard" Question

Question 5

Question 5

For each area (renamed to country), generate a column that lists out the distinct items that the area produces. Order the results alphabetically. HINT: STRING_AGG()

Question 5 - Solution

Question 5

For each area (renamed to country), generate a column that lists out the distinct items that the area produces. Order the results alphabetically. HINT: STRING_AGG()

Possible Solution:

```
SELECT area AS country,
    STRING_AGG(DISTINCT item, '  -  ') AS produce
FROM crop_livestock_stats
GROUP BY country
ORDER BY country;
```

Summary

- We practiced SQL through exercises using the FAOSTAT dataset
- You can reference the possible solutions in these slides and document here, which will be shared after class

Thank You

And Keep Practicing!