Chapter 4: Complex Calculations

Exercise 3: Most decorated athlete per region

Your goal for this exercise is to show the most decorated athlete per region. To set up this report, you need to leverage the ROW_NUMBER() window function, which numbers each row as an incremental integer, where the first row is 1, the second is 2, and so on.

Syntax for this window function is ROW_NUMBER() OVER (PARTITION BY field ORDER BY field). Notice how there is no argument within the initial function.

When set up correctly, a row_num = 1 represents the most decorated athlete within that region. Note that you cannot use a window calculation within a HAVING or WHERE statement, so you will need to use a subquery to filter.

Feel free to reference the E:R Diagram. We will use summer_games_clean to avoid null handling.

Instructions for the exercises:

- Query 1:
    - Build a query that pulls region, athlete_name, and total_golds by joining summer_games_clean, athletes, and countries.
    - Add a field called row_num that uses ROW_NUMBER() to assign a regional rank to each athlete based on total golds won.
- Query 2:
    - Alias the subquery as subquery
    - Query region, athlete_name, and total_golds, and then filter for only the top athlete per region.

Sample database:

query result | summer_games | countries | country_stats | **athletes**

id	name	gender	age	height	weight
51	Nstor Abad Sanjun	M	23	167	64
55	Antonio Abadia Beci	M	26	170	65
110	Abubakar Abbas Abbas	M	20	175	66
126	Forough Abbasi	F	20	164	58
251	Bashir Abdi	M	27	176	56
273	Maizurah Abdul Rahim	F	17	147	50

Showing 100 out of 4216 rows

query result | **summer_games_clean** | countries

sport	event	year	athlete_id	country_id	bronze	silver	gold
Gymnastics	Gymnastics Men's Individual All-Around	2016-01-01	51	173	0	0	0
Gymnastics	Gymnastics Men's Floor Exercise	2016-01-01	51	173	0	0	0
Gymnastics	Gymnastics Men's Parallel Bars	2016-01-01	51	173	0	0	0
Gymnastics	Gymnastics Men's Horizontal Bar	2016-01-01	51	173	0	0	0
Gymnastics	Gymnastics Men's Rings	2016-01-01	51	173	0	0	0

Showing 100 out of 4937 rows

query result | summer_games | **countries** | country_stats | athletes

id	country	region
1	AFG - Afghanistan	ASIA (EX. NEAR EAST)
2	ALB - Albania	EASTERN EUROPE
3	ALG - Algeria	NORTHERN AFRICA
4	A.Sa - American Samoa	OCEANIA
5	AND - Andorra	WESTERN EUROPE
6	ANG - Angola	SUB-SAHARAN AFRICA

Showing 100 out of 203 rows

Query given as the solution:

Query 1:

```
SELECT
        -- Query region, country, and total gold medals
        region,
    name AS athlete_name,
    SUM(gold) AS total_golds,
    -- Assign a regional rank to each athlete
    ROW_NUMBER() OVER (PARTITION by region ORDER BY SUM(gold) DESC) AS row_num
FROM summer_games_clean
JOIN countries
ON summer_games_clean.country_id = countries.id
JOIN athletes
ON summer_games_clean.athlete_id = athletes.id
GROUP BY region, athlete_name;
```

Query 2:

```
-- Query region, athlete name, and total_golds
SELECT
        region,
    athlete_name,
    total_golds
FROM
    (SELECT
        -- Query region, country, and total gold medals
        region,
        name AS athlete_name,
        SUM(gold) AS total_golds,
        -- Assign a regional rank to each athlete
        ROW_NUMBER() OVER (PARTITION BY region ORDER BY SUM(gold) DESC) AS row_num
    FROM summer_games_clean AS s
    JOIN athletes AS a
    ON a.id = s.athlete_id
    JOIN countries AS c
    ON s.country_id = c.id
    -- Alias as subquery
    GROUP BY region, athlete_name) AS subquery
-- Filter for only the top athlete per region
WHERE row_num = 1;
```