Suppose you’re an analyst for an e-commerce store. You’re trying to identify the top selling products in Q4 2017 by region, and you have 2 tables that you can query:

Table: all\_products

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| product\_id | integer | id of the product |
| product\_name | string | name of the product |
| sku | integer | universal stock keeping unit number |
| distributor\_id | integer | id for distributor |

Table: orders

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| date | string | format is "YYYY-MM-DD" |
| user\_id | integer | id of purchaser |
| order\_id | integer | id of order number |
| product\_id | integer | id of product |
| no\_units | integer | number of units sold in the order |
| price | integer | price per item |
| shipping\_id | integer | id of shipment |
| region | string | region being shipped to |

Using the tables above, write a SQL query to find the **top 5 selling products (in terms of total units sold) by region in Q4 of 2017**. Include both the distributor id as well as the name of the product in your results.

**My Answer:**

SELECT

b.region,

a.distributor\_id,

a.product\_name,

b.num\_units\_sold

FROM

(SELECT

distributor\_id,

product\_id,

product\_name

FROM all\_products) a

JOIN

(

SELECT

Region,

Product\_id,

Num\_units\_sold,

RANK() OVER (ORDER BY num\_units\_sold)as ranking

FROM

(

SELECT

region,

product\_id,

sum(no\_units) as num\_units\_sold

FROM orders

WHERE date>= "2017-10-01" and date <"2018-01-01"

GROUP BY region, product\_id)

) b

ON (a.product\_id=b.product\_id)

Where a.ranking<=5

ORDER BY a.region, b.num\_units\_sold desc