4️⃣

Highest-Grossing Items (Amazon)

QUESTION

Assume you're given a table containing data on Amazon customers and their spending on products in different categorie, write a query to identify the top two highest-grossing products within each category in the year 2022. The output should include the category, product, and total spend.

`product_spend` Table:

Column Name	Type
category	string
product	string
user_id	integer
spend	decimal
transaction_date	timestamp

`product_spend` Example Input:

category	product	user_id	spend	transaction_date
appliance	refrigerator	165	246.00	12/26/2021 12:00:00
appliance	refrigerator	123	299.99	03/02/2022 12:00:00
appliance	washing machine	123	219.80	03/02/2022 12:00:00
electronics	vacuum	178	152.00	04/05/2022 12:00:00
electronics	wireless headset	156	249.90	07/08/2022 12:00:00
electronics	vacuum	145	189.00	07/15/2022 12:00:00

Step 1: Identify the problem of the case

CASE offers a database with data on Amazon users and their spending on several categories of items; create a query to discover the two most earning products in each category in 2022. Outputs should comprise categories, goods, and total expenditures, and they should have the following form:

Column Name	Type
category	string
product	string
total_spend	demical

Step 2 : Analyze and solve problems

To determine the highest-grossing items, we must first determine the overall expenditure for each category and product. We could additionally limit the transactions to those from the year 2022.

```
SELECT category, product, SUM(spend) total_spend
FROM product_spend
WHERE EXTRACT(year FROM transaction_date) = 2022
GROUP BY category, product
```

We may use the RANK() window function inside the same query to get the ranking of goods based on total spent within each category.

```
SELECT category, product, SUM(spend) total_spend
      , RANK() OVER(PARTITION BY category ORDER BY SUM(spend) DESC)
FROM product_spend
WHERE EXTRACT(year FROM transaction_date) = 2022
GROUP BY category, product
```

We may utilize the preceding query as a CTE and filter for products with a ranking less than or equal to 2 because we're only interested in the top two highest-grossing products.

```
SELECT category,  product,  total_spend
FROM (
SELECT category, product, SUM(spend) total_spend
      , RANK() OVER(PARTITION BY category ORDER BY SUM(spend) DESC) rank_spend
FROM product_spend
WHERE EXTRACT(year FROM transaction_date) = 2022
GROUP BY category, product) sub
WHERE rank_spend <= 2
```

We have the final output:

category	product	total_spend
appliance	washing machine	439.80
appliance	refrigerator	299.99
electronics	vacuum	486.66
electronics	wireless headset	447.90