
3.7: Joining Tables of Data

Innocent Bayai

12 June 2024

1. Write a query to find the top 10 countries for Rockbuster in terms of customer numbers. (Tip: you'll have to use **GROUP BY** and **ORDER BY**, both of which follow the join.)

Thinking process

The country table is linked to customer table through address and city tables, i.e. customer to address to city to country. The tables are given alphabetic notations as follows:

A. customer, B. address, C. city, and D. country

The keys that link these tables are as follows:

B on A address_id

C on B city_id

D on C country_id

But what is required is just the top 10 countries i.e. LIMIT 10 in SQL, and the number of customers in each country i.e COUNT (customer_id) given alias AS customer_numbers. The customer_numbers must be GROUPED BY country and presented in descending order, i.e. ORDER BY DESC

Based on this the SQL query is as follows:

```
SELECT D. country
       COUNT (DISTINCT customer_id) AS customer_numbers
FROM customer A.
INNER JOIN address B ON A. address_id = B. address_id
INNER JOIN city C ON B. city_id = C. city_id
INNER JOIN country D ON C. country_id = D. country_id
GROUP BY country
ORDER BY customer_numbers DESC
```

The output is as follows:

```

50 SELECT D. country,
51        COUNT (DISTINCT customer_id) AS customer_numbers
52 FROM customer A
53 INNER JOIN address B ON A. address_id = B. address_id
54 INNER JOIN city C ON B. city_id = C. city_id
55 INNER JOIN country D ON C. country_id = D. country_id
56 GROUP BY country
57 ORDER BY customer_numbers DESC
58 LIMIT 10

```

Data Output Messages Notifications

	country character varying (50)	customer_numbers bigint
1	United States	33
4	Japan	31
5	Mexico	30
6	Brazil	28
7	Russian Federation	28
8	Philippines	20
9	Turkey	15
10	Indonesia	14

Total rows: 10 of 10 Query complete 00:00:00.078

- Next, write a query to identify the top 10 cities that fall within the top 10 countries you identified in step 1. (Hint: the top 10 cities can be in any of the countries identified—you don't need to create a separate list for each country.)

Write a short explanation of how you approached this query and why.

Thinking process

I need to show country in one column, followed by city column and the column providing a COUNT (customer_id) AS customer_numbers.

The tables are connected and identified by alphabetical letters as follows: A. customer, B. address, C. city, and D. country. As in the prior question, the keys linking these tables are as follows:

B on A address_id
 C on B city_id
 D on C country_id

The new query builds on top of the prior query to get the top 10 cities out of the top 10 countries already identified in the prior question. Filtering for country, city and descending order is achieved through ORDER BY and GROUP BY.

Query:

```
SELECT D. country,
       C. city
       COUNT (DISTINCT customer_id) AS customer_numbers
FROM customer A
JOIN address B ON A. address_id = B. address_id
JOIN city C ON B. city_id = C. city_id
JOIN country D ON C. country_id = D. country_id
WHERE
D. country IN (
SELECT country
FROM customer A
JOIN address B ON A. address_id = B. address_id
JOIN city C ON B. city_id = C. city_id
JOIN country D ON C. country_id = D. country_id
GROUP BY D. country
ORDER BY COUNT(A.customer_id) DESC
LIMIT 10
)
GROUP BY D.country, C. city
ORDER BY customer_numbers DESC
LIMIT 10;
```

The output is as follows.

60	SELECT D. country,
61	C. city,
62	COUNT (A.customer_id) AS customer_numbers
63	FROM customer A
64	JOIN address B ON A. address_id = B. address_id
65	JOIN city C ON B. city_id = C. city_id
66	JOIN country D ON C. country_id = D. country_id
67	WHERE
68	D. country IN (
69	SELECT country
70	FROM customer A
71	JOIN address B ON A. address_id = B. address_id
72	JOIN city C ON B. city_id = C. city_id
73	JOIN country D ON C. country_id = D. country_id
74	GROUP BY D. country
75	ORDER BY COUNT(A.customer_id) DESC
76	LIMIT 10
77)
78	GROUP BY D.country, C. city
79	ORDER BY customer_numbers DESC
80	LIMIT 10

Data Output	Messages	Notifications
-------------	----------	---------------

country	city	customer_numbers
character varying (50)	character varying (50)	bigint
8	Russian Federation	Teboksary
9	China	Tianjin
10	Indonesia	Cianjur

Total rows: 10 of 10	Query complete 00:00:00.070
----------------------	-----------------------------

- Now write a query to find the top 5 customers from the top 10 cities who've paid the highest total amounts to Rockbuster. The customer team would like to reward them for their loyalty!

Tip: After the join syntax, you'll need to use the WHERE clause with an operator, followed by GROUP BY and ORDER BY. Your output should include the following columns: Customer ID, Customer First Name and Last Name, Country, City, and Total Amount Paid.

Thinking process

The right columns (payment, customer-id, first_name, last_name, country and city) are picked from the tables linked to customer, address, city and country whilst observing the SUM of payments made each customer. Joins are made given primary and foreign keys that connect the tables. A subquery is adopted to cater for the countries and cities that have the highest total payments made by customers. Ordering and grouping are used to give the output the required format. The five tables are identified as follows:

A. payment, B. customer, C. address, D. city, E. country

The tables are connected by the following keys:

A and B – customer_id

B and C – address_id

C and D – city_id

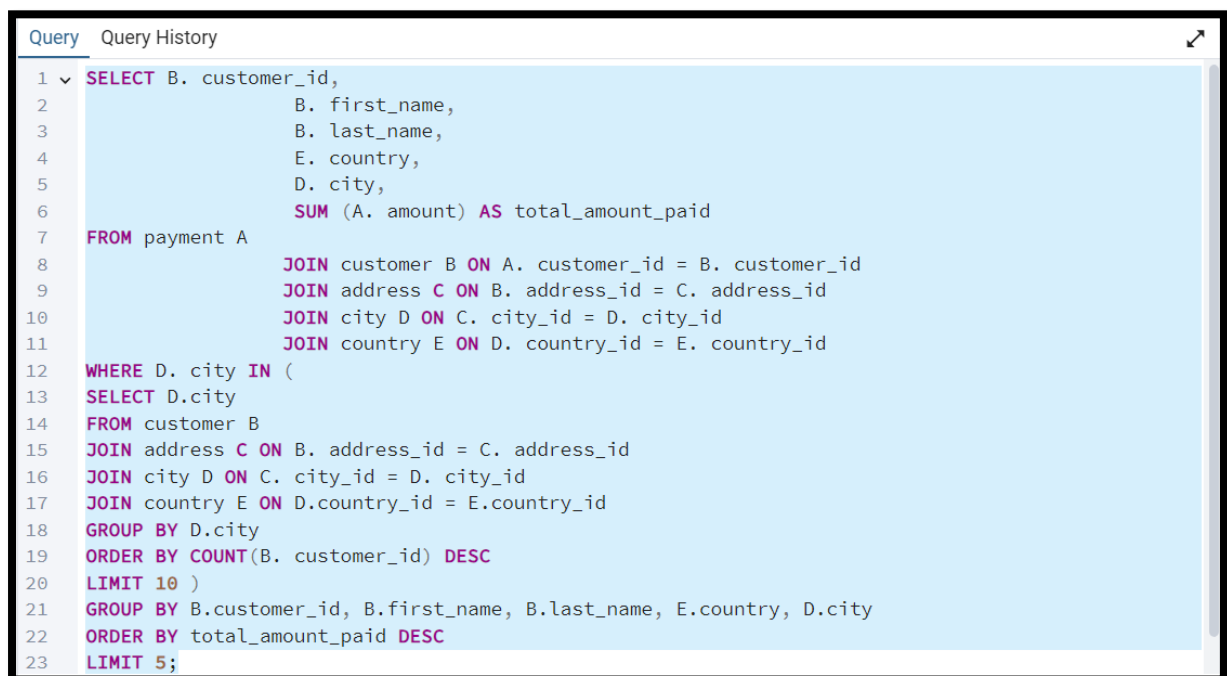
D and E - country_id

Query is as follows:

```

SELECT B. customer_id,
       B. first_name,
       B. last_name,
       E. country,
       D. city,
       SUM (A. amount) AS total_amount_paid
FROM payment A
      JOIN customer B ON A. customer_id = B. customer_id
      JOIN address C ON B. address_id = C. address_id
      JOIN city D ON C. city_id = D. city_id
      JOIN country E ON D. country_id = E. country_id
WHERE
D. city IN (
SELECT D.city
FROM customer B
JOIN address C ON B. address_id = C. address_id
JOIN city D ON C. city_id = D. city_id
JOIN country E ON D.country_id = E.country_id
GROUP BY D.city
ORDER BY COUNT(B. customer_id) DESC
LIMIT 10
)
GROUP BY B.customer_id, B.first_name, B.last_name, E.country, D.city
ORDER BY total_amount_paid DESC
LIMIT 5;

```



```

1  SELECT B. customer_id,
2      B. first_name,
3      B. last_name,
4      E. country,
5      D. city,
6      SUM (A. amount) AS total_amount_paid
7  FROM payment A
8      JOIN customer B ON A. customer_id = B. customer_id
9      JOIN address C ON B. address_id = C. address_id
10     JOIN city D ON C. city_id = D. city_id
11     JOIN country E ON D. country_id = E. country_id
12  WHERE D. city IN (
13  SELECT D.city
14  FROM customer B
15  JOIN address C ON B. address_id = C. address_id
16  JOIN city D ON C. city_id = D. city_id
17  JOIN country E ON D.country_id = E.country_id
18  GROUP BY D.city
19  ORDER BY COUNT(B. customer_id) DESC
20  LIMIT 10 )
21  GROUP BY B.customer_id, B.first_name, B.last_name, E.country, D.city
22  ORDER BY total_amount_paid DESC
23  LIMIT 5;

```

Data Output							Messages	Notifications
<div><div><div><div></div><div></div><div></div><div></div><div></div><div></div><div></div><div></div></div></div></div>								
	customer_id integer	first_name character varying (45)	last_name character varying (45)	country character varying (50)	city character varying (50)	total_amount_paid numeric		
1	148	Eleanor	Hunt	Runion	Saint-Denis	211.55		
2	144	Clara	Shaw	Belarus	Molodetno	189.60		
3	566	Casey	Mena	Turkey	Tokat	130.68		
4	84	Sara	Perry	Mexico	Atlixco	128.70		
5	506	Leslie	Seward	Indonesia	Pontianak	123.72		
Total rows: 5 of 5							Query complete 00:00:00.089	