

Data Analytics Immersion

3.7: Joining Tables of Data

Mehreen Becker

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.)
 - o Copy-paste your query and its output into your answers document.

The screenshot shows the pgAdmin 4 interface. On the left is the Object Explorer tree, which includes a 'Tables (16)' node under 'Rockbuster/postgres@My Server'. The main area contains a SQL editor window with the following query:

```
1 SELECT country AS country_name,
2 COUNT (customer_id) AS customer_count
3 FROM customer
4 JOIN address ON customer.address_id = address.address_id
5 JOIN city ON address.city_id = city.city_id
6 JOIN country ON city.country_id = country.country_id
7 GROUP BY country.country
8 ORDER BY COUNT(customer_id) DESC
9 LIMIT 10;
```

Below the SQL editor is a Data Output grid showing the results of the query:

country_name	customer_count
India	60
China	53
United States	36
Japan	31
Mexico	30
Brazil	28
Russian Federation	28
Philippines	20
Turkey	15
Indonesia	14

Total rows: 10 Query complete 00:00:00.076

- o Write a few sentences on how you approached this query and why. You must be able to explain your thought process when writing queries, especially for future interviews. I began by identifying the tables that store customer and location data (customer, address, city and country) and determined how they related through their foreign keys. Then, I joined these tables to connect each customer to their corresponding country and used COUNT to calculate the total number of customers per country. Finally, I sorted the results in descending order and limited the output to 10 countries as requested.
2. 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.)
 - o Copy-paste your query and its output into your answers document.

The screenshot shows the pgAdmin 4 interface. In the Object Explorer, under Schemas (1) > public > Tables (16), the 'customer' table is selected. The main window displays a SQL query in the Query tab:

```

1 SELECT city, country,
2     COUNT (customer_id) AS customer_count
3 FROM customer
4 JOIN address AS address ON customer.address_id = address.address_id
5 JOIN city AS city ON address.city_id = city.city_id
6 JOIN country AS country ON city.country_id = country.country_id
7 WHERE country IN ('India','China','United States','Japan','Mexico','Brazil',
8 'Russian Federation','Philippines','Turkey','Indonesia')
9 GROUP BY city, country
10 ORDER BY customer_count DESC
11 LIMIT 10;

```

The Data Output tab shows the results of the query:

	city	country	customer_count
1	Aurora	United States	2
2	Atilxco	Mexico	1
3	Xintai	China	1
4	Adoni	India	1
5	Dhule (Dhulia)	India	1
6	Kurashiki	Japan	1
7	Pingxiang	China	1
8	Sivas	Turkey	1
9	Celaya	Mexico	1
10	So Leopoldo	Brazil	1

Total rows: 10 Query complete 00:00:00.062 LF Ln 11, Col 10

- Write a short explanation of how you approached this query and why. The steps remained the same for the most part as in step 1. Since we had already identified the top 10 countries, I looked at the cities in those countries by looking at the customer count. Finally, I sorted the results in descending order and limited the output to 10 countries as requested.
3. 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.
 - Copy-paste your query and its output into your answers document.

pgAdmin 4 Object Tools Edit View Window Help pgAdmin 4 Mon 20 Oct 12:00

Object Explorer Processes Testing/postgres... Testing/postgres... postgres/postgres... Preferences Rockbuster/postgres@My Server*

Query History Scratch Pad

```
1 SELECT customer.customer_id, customer.first_name, customer.last_name, country.country, city.city,
2 SUM (payment.amount) AS total_payment
3 FROM customer AS customer
4 JOIN payment AS payment ON customer.customer_id = payment.customer_id
5 JOIN address AS address ON customer.address_id = address.address_id
6 JOIN city AS city ON address.city_id = city.city_id
7 JOIN country AS country ON city.country_id = country.country_id
8 WHERE city IN ('Aurora', 'Atlixco', 'Xintai', 'Adoni', 'Dhule (Dhulia)', 'Kurashiki', 'Pingxiang', 'Sivas',
9 AND country IN ('India', 'China', 'United States', 'Japan', 'Mexico', 'Brazil', 'Russian Federation', 'Philippines')
10 GROUP BY customer.customer_id, customer.first_name, customer.last_name, country.country, city.city
11 ORDER BY total_payment DESC
12 LIMIT 5;
```

Data Output Messages Notifications

Showing rows: 1 to 5 Page No: 1 of 1

	customer_id	first_name	last_name	country	city	total_payment
1	84	Sara	Perry	Mexico	Atlixco	128.70
2	518	Gabriel	Harder	Turkey	Sivas	108.75
3	587	Sergio	Stanfield	Mexico	Celaya	102.76
4	537	Clinton	Buford	United States	Aurora	98.76
5	367	Adam	Gooch	India	Adoni	97.80

Total rows: 5 Query complete 00:00:00.068 LF Ln 9, Col 4

The screenshot shows the pgAdmin 4 interface. In the top navigation bar, it says 'pgAdmin 4' and 'Mon 20 Oct 12:00'. Below the bar are tabs for 'Object Explorer', 'Processes', 'Testing/postgres...', 'postgres/postgres...', 'Preferences', and 'Rockbuster/postgres@My Server*'. The 'Object Explorer' tab is selected. On the left, there's a tree view of database objects under 'Tables (16)' which includes Actor, address, category, city, country, customer, film, film_actor, film_category, inventory, language, payment, rental, staff, store, Trigger Functions, Types, and Views. The main area contains a SQL editor with a multi-line query. Below the editor is a 'Data Output' panel showing the results of the query as a table. The table has columns: customer_id, first_name, last_name, country, city, and total_payment. It lists five rows of data. At the bottom of the Data Output panel, it says 'Total rows: 5 Query complete 00:00:00.068'. The status bar at the bottom right shows 'LF Ln 9, Col 4'.