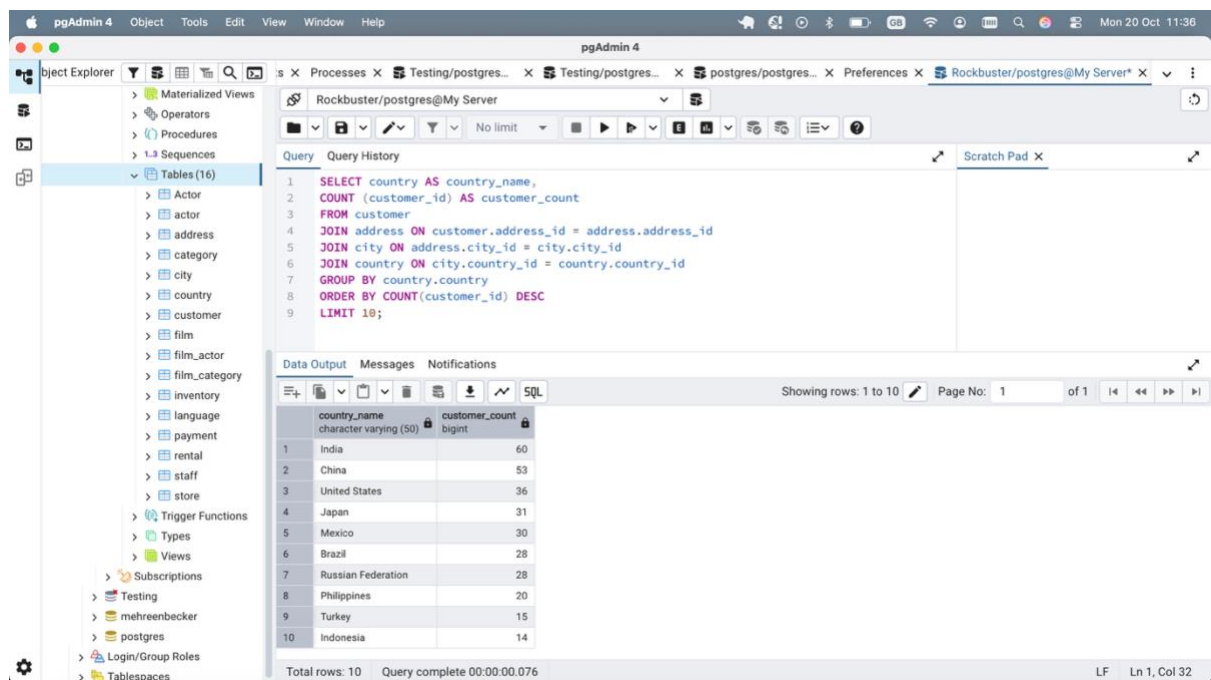


Data Analytics Immersion

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



The screenshot shows the pgAdmin 4 interface. On the left, the Object Explorer shows the database structure. The main pane displays a SQL query in the Query editor. The query is as follows:

```
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;
```

The Data Output pane shows the results of the query, displaying the top 10 countries by customer count:

	country_name	customer_count
1	India	60
2	China	53
3	United States	36
4	Japan	31
5	Mexico	30
6	Brazil	28
7	Russian Federation	28
8	Philippines	20
9	Turkey	15
10	Indonesia	14

The status bar at the bottom indicates "Total rows: 10" and "Query complete 00:00:00.076".

- 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.)
 - 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 with a tree view of the database schema. The main pane displays a SQL query and its results. The query is as follows:

```

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 results pane shows a table with 10 rows. The columns are city, country, and customer_count. The data is as follows:

city	country	customer_count
Aurora	United States	2
Atlisco	Mexico	1
Xintai	China	1
Adoni	India	1
Dhule (Dhulia)	India	1
Kurashiki	Japan	1
Pingxiang	China	1
Sivas	Turkey	1
Celaya	Mexico	1
So Leopoldo	Brazil	1

The status bar at the bottom indicates "Total rows: 10" and "Query complete 00:00:00.062".

- 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 started 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

Rockbuster/postgres@My Server

Query Query History

```
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', 'Si
9 AND country IN ('India','China','United States','Japan', 'Mexico','Brazil', 'Russian Federation','Phil
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 integer	first_name character varying (45)	last_name character varying (45)	country character varying (50)	city character varying (50)	total_payment numeric
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