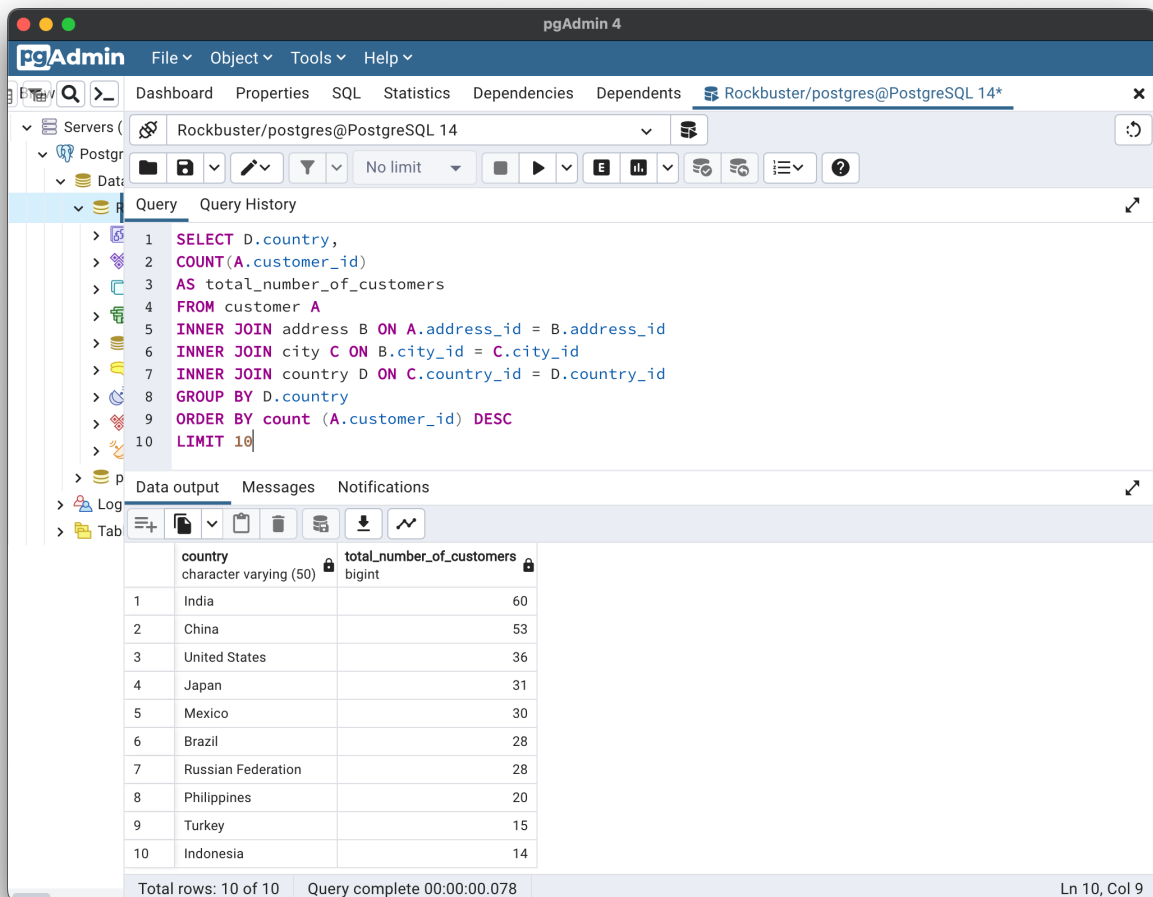


3.7: Joining Tables of Data

Alejandro Salgado

Step 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.)



The screenshot shows the pgAdmin 4 interface. The SQL query editor contains the following query:

```
1 SELECT D.country,  
2 COUNT(A.customer_id)  
3 AS total_number_of_customers  
4 FROM customer A  
5 INNER JOIN address B ON A.address_id = B.address_id  
6 INNER JOIN city C ON B.city_id = C.city_id  
7 INNER JOIN country D ON C.country_id = D.country_id  
8 GROUP BY D.country  
9 ORDER BY count (A.customer_id) DESC  
10 LIMIT 10
```

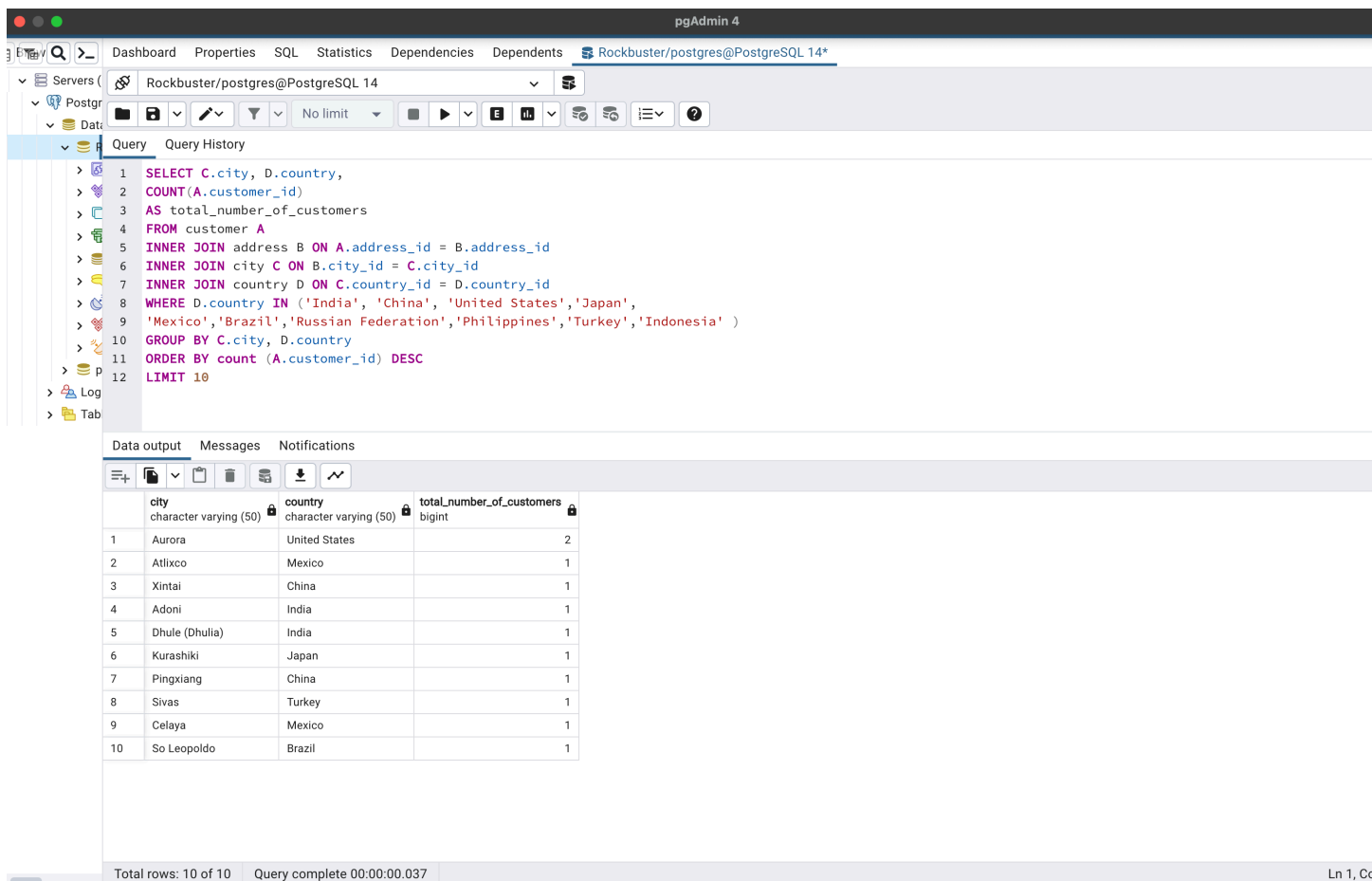
The query results are displayed in a table with two columns: **country** (character varying (50)) and **total_number_of_customers** (bigint). The results show the top 10 countries by customer count.

	country	total_number_of_customers
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

Total rows: 10 of 10 | Query complete 00:00:00.078 | Ln 10, Col 9

- The first step I did was look at the ERD table and see what the connections were. I saw that address id, city id and country id were connected. I figured INNER JOIN would be the best to use. I used SELECT for the country which was table D and then used COUNT since we are going to count the number of customers. Used AS for the column name and then started joining the tables. After that I used the GROUP BY command to sum up each country and ORDER BY to see the order by customers in descending order and since we are only seeing the top 10, there was a LIMIT of 10.

Step 2: Write a query to find the top 10 cities within the top 10 countries identified in step 1.



The screenshot shows the pgAdmin 4 interface. The SQL query editor contains the following query:

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

The query results are displayed in the Data output tab, showing 10 rows of data:

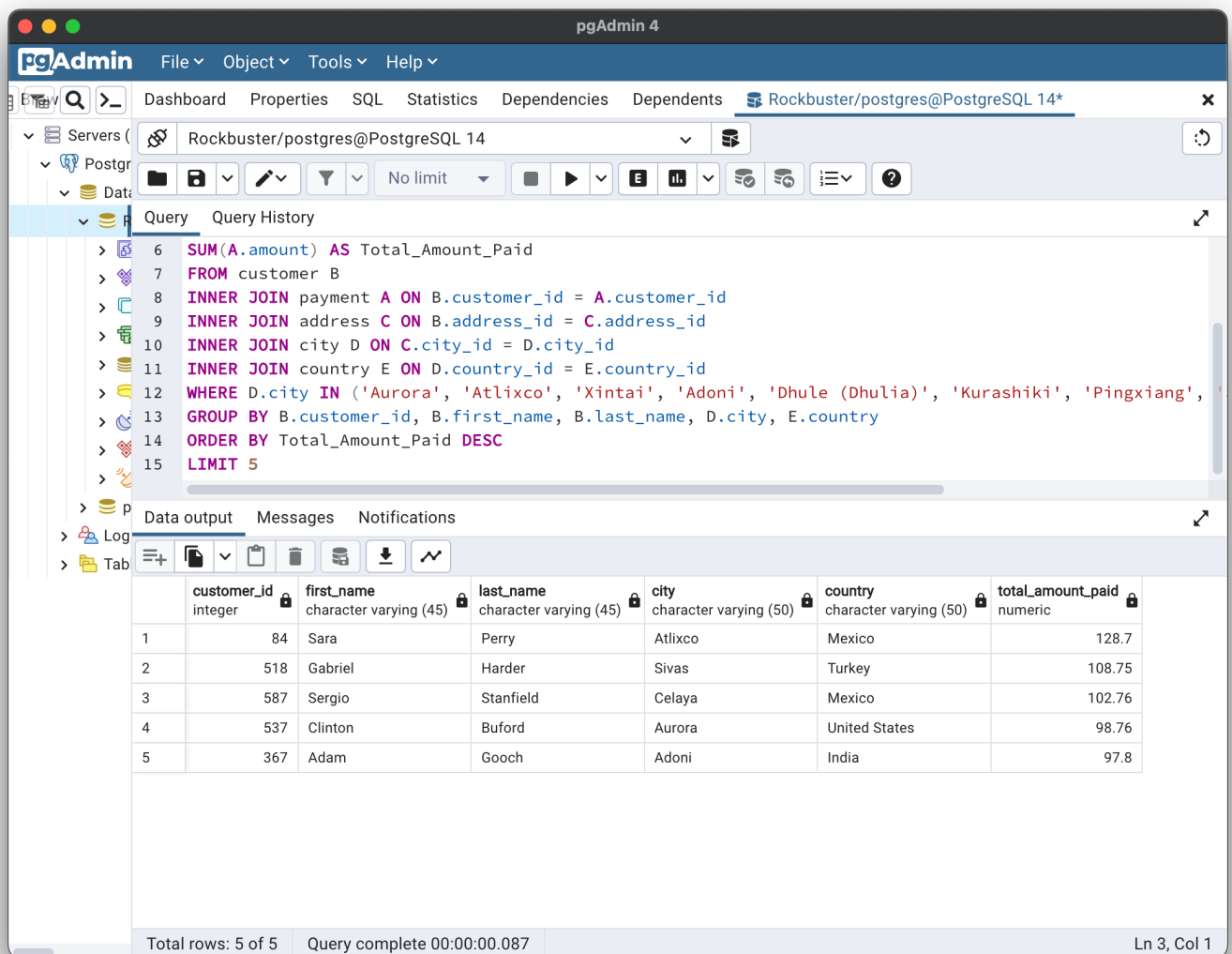
	city character varying (50)	country character varying (50)	total_number_of_customers bigint
1	Aurora	United States	2
2	Atlixco	Mexico	1
3	Xintai	China	1
4	Adoni	India	1
5	Dhule (Dhulia)	India	1
6	Kurashiki	Japan	1
7	Pingxiang	China	1
8	Sivas	Turkey	1
9	Celaya	Mexico	1
10	So Leopoldo	Brazil	1

Total rows: 10 of 10 Query complete 00:00:00.037 Ln 1, Co

- For this query I found that it is similar to the last one. I added the city in the SELECT command because we are trying to look for the top 10 cities within the top 10 countries. I kept the same INNER JOIN commands, but when it comes to where, I added the top 10 countries from the last query and I added the city to the GROUP BY command. Just like the last query, I put a LIMIT of 10.

Step 3: Write a query to find the top 5 customers in the top 10 cities who have 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, Total Amount Paid.



The screenshot shows the pgAdmin 4 interface with a SQL query executed. The query is as follows:

```
6 SUM(A.amount) AS Total_Amount_Paid
7 FROM customer B
8 INNER JOIN payment A ON B.customer_id = A.customer_id
9 INNER JOIN address C ON B.address_id = C.address_id
10 INNER JOIN city D ON C.city_id = D.city_id
11 INNER JOIN country E ON D.country_id = E.country_id
12 WHERE D.city IN ('Aurora', 'Atlixco', 'Xintai', 'Adoni', 'Dhule (Dhulia)', 'Kurashiki', 'Pingxiang',
13 GROUP BY B.customer_id, B.first_name, B.last_name, D.city, E.country
14 ORDER BY Total_Amount_Paid DESC
15 LIMIT 5
```

The results are displayed in a table with the following columns: customer_id, first_name, last_name, city, country, and total_amount_paid. The table shows 5 rows of data.

customer_id	first_name	last_name	city	country	total_amount_paid
84	Sara	Perry	Atlixco	Mexico	128.7
518	Gabriel	Harder	Sivas	Turkey	108.75
587	Sergio	Stanfield	Celaya	Mexico	102.76
537	Clinton	Buford	Aurora	United States	98.76
367	Adam	Gooch	Adoni	India	97.8

Total rows: 5 of 5 Query complete 00:00:00.087 Ln 3, Col 1