

## 3.7 Joining Tables of Data

Rockbuster's management team would like to know the top 10 countries where Rockbuster customers are based so they can focus on building a better brand image in those markets.

1. **Check for and clean dirty data: 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.

```
SELECT co.country AS customer_country,  
       COUNT(cu.customer_id) AS number_of_customer  
FROM customer AS cu  
INNER JOIN address AS ad ON cu.address_id = ad.address_id  
INNER JOIN city AS ci ON ad.city_id = ci.city_id  
INNER JOIN country AS co ON ci.country_id = co.country_id  
GROUP BY country  
ORDER BY number_of_customer DESC  
LIMIT 10;
```








Data output Messages Notifications		
	customer_country character varying (50)	number_of_customer bigint
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 00:00:00.107

- Write a few sentences on how you approached this query and why. It's important that you can explain your thought process when writing queries, especially for future interviews.
  1. Firstly, I identify what final information I want to get from the query.
  2. From the ERD, I identify which table could provide me that final information from step 1.
  3. Once done, it starts with the question on how? Do I only need one table or multiple tables?  
In case multiple tables:
    - based on step 1 and 2, identify the joins (INNER JOIN, LEFT/RIGHT JOIN, FULL JOIN)
    - identify the path to the desired table
    - list the required table with their common keys.
  4. Write the query from step 1-3 including the additional necessary filters (WHERE, GROUP BY, ORDER BY, LIMIT, etc.)

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

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

```
SELECT ci.city AS customer_city,  
       co.country AS customer_country,  
       COUNT(cu.customer_id) AS number_of_customer  
FROM customer AS cu  
INNER JOIN address AS ad ON cu.address_id = ad.address_id  
INNER JOIN city AS ci ON ad.city_id = ci.city_id  
INNER JOIN country AS co ON ci.country_id = co.country_id  
WHERE country IN ('India', 'China', 'United States', 'Japan', 'Mexico',  
                  'Brazil', 'Russian Federation', 'Philippines', 'Turkey', 'Indonesia')  
GROUP BY city, country  
ORDER BY number_of_customer DESC  
LIMIT 10;
```

Data output Messages Notifications			
      			
	customer_city character varying (50)	customer_country character varying (50)	number_of_customer 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 00:00:00.126	

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

I always identify the query based on the desired final information. The query “to find the top 10 cities within the top 10 countries identified in step 1” is indeed a continuation of the query done in number 1. Here, I only need to go one layer deeper, which is: city from those country. Thus, WHERE is used.

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

```
SELECT cu.customer_id,  
       cu.first_name,  
       cu.last_name,  
       ci.city,  
       co.country,  
       SUM(pay.amount) AS total_amount_paid  
FROM customer as cu  
INNER JOIN address AS ad ON cu.address_id = ad.address_id  
INNER JOIN city AS ci ON ad.city_id = ci.city_id  
INNER JOIN country AS co ON ci.country_id = co.country_id  
INNER JOIN payment AS pay ON cu.customer_id = pay.customer_id  
WHERE city IN ('Aurora', 'Atlixco', 'Xintai', 'Adoni', 'Dhule(Dhulia)', 'Kurashiki',  
              'Pingxiang', 'Sivas', 'Celaya', 'So Leopoldo')  
GROUP BY cu.customer_id, first_name, last_name, city, country  
ORDER BY total_amount_paid DESC  
LIMIT 5;
```

Data output Messages Notifications						
	customer_id integer	first_name character varying (45)	last_name character varying (45)	city character varying (50)	country character varying (50)	total_amount_paid numeric
1	84	Sara	Perry	Atlixco	Mexico	128.70
2	518	Gabriel	Harder	Sivas	Turkey	108.75
3	587	Sergio	Stanfield	Celaya	Mexico	102.76
4	537	Clinton	Buford	Aurora	United States	98.76
5	367	Adam	Gooch	Adoni	India	97.80