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.**

Query | Query History

```
SELECT D.country,
       COUNT(A.customer_id) AS customer_numbers_count
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_count DESC
LIMIT 10;
```

Data Output | Messages | Notifications

	country character varying (50)	customer_numbers_count 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

- **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.**

  The manager at Rockbuster Stealth wants to find the top 10 countries for Rockbuster in terms of customer numbers. The first step is to find out which tables in the Rockbuster database hold the data we need. Based on the ERD, columns “customer_id” and “country” are stored in the “customer” and “country” table.
  In the SELECT clause, the customer_id and country column will be used to find the top 10 countries for Rockbuster in terms of customer numbers. We are also looking for the customer numbers and will use the aggregate function COUNT. Additionally, an alias will be given for the customer_id column to avoid ambiguity.

Each column is marked A, B, C, or D to show which table it appears in. The customer table is A, address table is B, city table is C, and country table is D.
Since it is not possible to join the “customer” table directly with the “country” table, we will **need to use INNER JOIN** to merge the customer table with the address table, the address table with the city table, and the city table with country table to get the relevant information. The **INNER JOIN** is the most suitable type of join for this query since this JOIN returns records that have a match in both tables based on the join predicate (which comes after the ON keyword).
I will use the customer table for **FROM clause** since the table after the FROM clause is always the left table. In this instance, the customer table is the left table and marked as table A for readability.
Again, we want to list the number of customers in each country, so we will also use the **GROUP BY clause** and include the country column. (As a reminder, when you are aggregating a variable in the SELECT clause you need every variable that is not the aggregated value in the GROUP BY clause)
We used the **ORDER BY clause** to sort the customer_id column in descending order to find out the top 10 countries for Rockbuster based on customer numbers.
Lastly, the **LIMIT 10 command** restricts the search to 10 rows to be pulled from the database.

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.**

Query | Query History

```
SELECT C.city,
       D.country,
       COUNT(A.customer_id) AS customer_numbers_count
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
WHERE country IN ('India','China','United States','Japan','Mexico','Brazil',
                  'Russian Federation','Philippines','Turkey','Indonesia')
GROUP BY city,
         country
ORDER BY customer_numbers_count DESC
LIMIT 10;
```

Data Output | Messages | Notifications

	city character varying (50)	country character varying (50)	customer_numbers_count 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

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

Since the question is asking to find the top 10 cities within the top 10 countries as identified in step 1, we will include columns “city,” “country” and “customer_numbers_count” (alias of customer_id) in the **SELECT clause**.
For the **WHERE clause**, we selected the country column as the condition to specify which rows we want, in this instance, we included the values of the top 10 countries from step 1 to identify the top 10 cities.
For the **GROUP BY clause**, we selected city and country since it is telling SQL what columns we want to use to aggregate the data (you want it to count the customers within) in order to find the top 10 cities within the top 10 countries.

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.**

FULL JOIN

Query | Query History | Scratch Pad

```
SELECT A.customer_id,
       B.first_name,
       B.last_name,
       D.city,
       E.country,
       SUM(A.amount) AS total_amount_paid
FROM payment A
FULL JOIN customer B ON A.customer_id = B.customer_id
FULL JOIN address C ON B.address_id = C.address_id
FULL JOIN city D ON C.city_id = D.city_id
FULL JOIN country E ON D.country_id = E.country_id
WHERE E.country IN ('India','China','United States','Japan','Mexico','Brazil',
                    'Russian Federation','Philippines','Turkey','Indonesia')
AND D.city IN ('Aurora','Atlixco','Xintai','Adoni','Dhule(Dhulia)','Kurashiki',
               'Pingxiang','Sivas','Celaya','So Leopoldo')
GROUP BY A.customer_id,
         B.first_name,
         B.last_name,
         D.city,
         E.country
ORDER BY total_amount_paid DESC
LIMIT 5;
```

Data Output | Messages | Notifications

	customer_id smallint	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

or

INNER JOIN

Query | Query History | Scratch Pad

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

Data Output | Messages | Notifications

	customer_id smallint	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