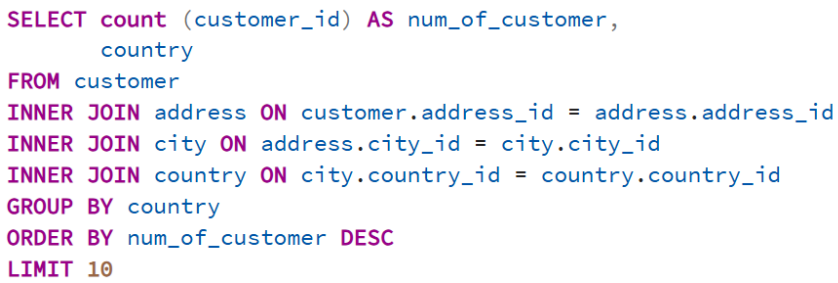
Joining Tables of Data

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.**
    - **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.**

Commands:

```
SELECT count(customer_id) AS num_of_customer,
        city,
        country
FROM customer
INNER JOIN address ON customer.address_id = address.address_id
INNER JOIN city ON address.city_id = city.city_id
INNER JOIN country ON city.country_id = country.country_id
GROUP BY city, country
ORDER BY num_of_customer DESC
LIMIT 10
```

Screenshots:



```
SELECT count (customer_id) AS num_of_customer,
       country
FROM customer
INNER JOIN address ON customer.address_id = address.address_id
INNER JOIN city ON address.city_id = city.city_id
INNER JOIN country ON city.country_id = country.country_id
GROUP BY country
ORDER BY num_of_customer DESC
LIMIT 10
```

	num_of_customer bigint	country character varying (50)
1	60	India
2	53	China
3	36	United States
4	31	Japan
5	30	Mexico
6	28	Brazil
7	28	Russian Federation
8	20	Philippines
9	15	Turkey
10	14	Indonesia

Explanations:

- Based on the ERD (Entity Relationship Diagram) from Exercise 3.2, the following tables are required to answer the question: **Customer → Address → City → Country**.
- An **INNER JOIN** is used because only the matching values between the Customer and Country tables (through Address and City) are required.
- **GROUP BY country** is applied to group the results by the Country column, as requested in the task.

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

Commands:

```
SELECT count(customer_id) AS num_of_customer,
        city,
        country
FROM customer
INNER JOIN address ON customer.address_id = address.address_id
INNER JOIN city ON address.city_id = city.city_id
INNER JOIN country ON city.country_id = country.country_id
WHERE country IN ('India', 'China', 'United States', 'Japan',
'Mexico', 'Brazil', 'Russian Federation', 'Philippines', 'Turkey', 'Indonesia')
GROUP BY city, country
ORDER BY num_of_customer DESC
```

```
LIMIT 10
```

Screenshots:

Query | Query History

```
SELECT count(customer_id) AS num_of_customer,
       city,
       country
FROM customer
INNER JOIN address ON customer.address_id = address.address_id
INNER JOIN city ON address.city_id = city.city_id
INNER JOIN country ON city.country_id = country.country_id
WHERE country in ('India', 'China', 'United States', 'Japan',
'Mexico', 'Brazil', 'Russian Federation', 'Philippines', 'Turkey', 'Indonesia'
)
GROUP BY city, country
ORDER BY num_of_customer DESC
Limit 10
```

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

Explanations:

- The **city** column is included in the output so that the results display the list of top 10 cities.
- A **WHERE** clause is applied to restrict the results to customers from the top 10 countries identified in Step 1.
- **GROUP BY city, country** ensures that the data is grouped first by city and then by country.
- **ORDER BY** is used to sort the output by the number of customers in descending order.
- **LIMIT 10** restricts the final output to only 10 rows.

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.**

Commands:

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

Screenshots:

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

	customer_id integer	first_name character varying (45)	last_name character varying (45)	country character varying (50)	city character varying (50)	total_amount 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

Explanations:

- Based on the ERD (Entity Relationship Diagram) from Exercise 3.2, the following tables are required to answer the question: **Payment → Customer → Address → City → Country**.
- An **INNER JOIN** is used because only the matching values between the Payment and Country tables (through Customer, Address and City) are required.
- A **WHERE** clause is applied to restrict the results to customers from the top 10 cities identified in Step 2.
- **GROUP BY cutomer_id, first_name, last_name, city, country** ensures that the data is grouped by the preferred order.
- **ORDER BY** is used to sort the output by the amount of payment in descending order.
- **LIMIT 5** restricts the final output to only 5 rows.