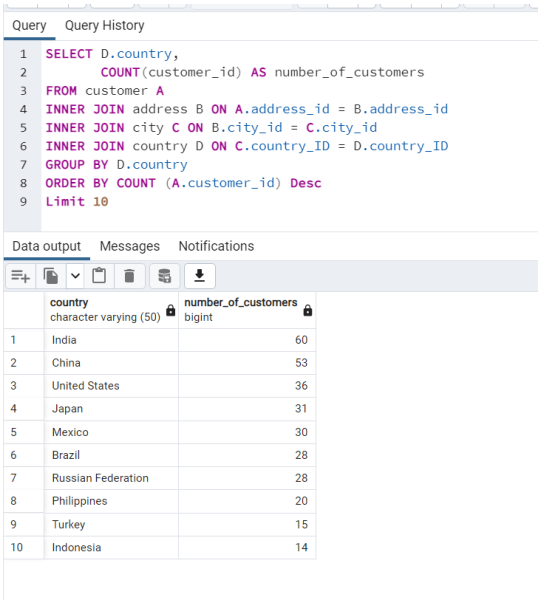
3.7 Joining Tables

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.)



Query | Query History

```
SELECT D.country,
       COUNT(customer_id) AS number_of_customers
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 D.country
ORDER BY COUNT (A.customer_id) Desc
Limit 10
```

Data output | Messages | Notifications

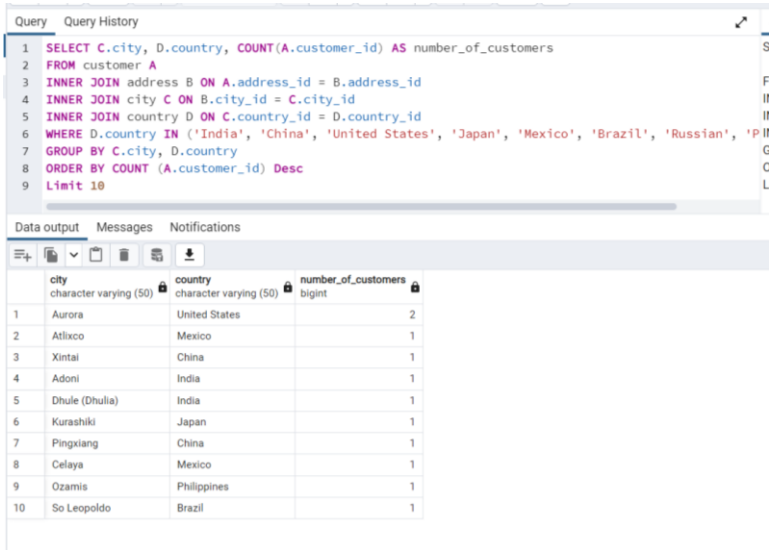
	country character varying (50)	number_of_customers bigint
1	India	60
2	China	53
3	United States	36
4	Japan	31
5	Mexico	30
6	Brazil	28
7	Russian Federation	28
8	Philippines	20
9	Turkey	15
10	Indonesia	14

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

- Steps from pulling the top 10 countries in term of customer numbers are
    - I created an ERD from the data dictionary, there are a list of tables and names of columns
    - The customer table and the country table have information on country and customer ID

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

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



Query | Query History

```
SELECT C.city, D.country, COUNT(A.customer_id) AS number_of_customers
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 D.country IN ('India', 'China', 'United States', 'Japan', 'Mexico', 'Brazil', 'Russian', 'P
GROUP BY C.city, D.country
ORDER BY COUNT (A.customer_id) Desc
Limit 10
```

Data output | Messages | Notifications

	city character varying (50)	country character varying (50)	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	Celaya	Mexico	1
9	Ozamis	Philippines	1
10	So Leopoldo	Brazil	1

Steps to write the query

- I checked the data dictionary front the ERD to know which tables to use.
- Like the first query but I added city variable in the select section of the query WHERE and IN syntax to restrict the list of countries to the top 10.
- Looking at the output there is only one city with more than one customer, the Aurora City in the United States has two customers. All the rest have only one.

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

Query | Query History

```
FROM customer A
INNER JOIN payment B ON A.customer_id = B.customer_id
INNER JOIN address C ON A.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', 'Atlixa', 'Xintai', 'Adoni', 'Dhule (Dhulia)',
                 'Kurashiki', 'Pingxang', 'Sivas', 'Celaya', 'So Leopoldo')
GROUP BY A.customer_id, A.first_name, A.last_name, D.city, E.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	518	Gabriel	Harder	Sivas	Turkey	108.75
2	587	Sergio	Stanfield	Celaya	Mexico	102.76
3	537	Clinton	Buford	Aurora	United States	98.76
4	367	Adam	Gooch	Adoni	India	97.8
5	443	Francisco	Skidmore	So Leopoldo	Brazil	93.79

Steps to write the query

In the third query we want to extract the top 5 customer in the top 10 cities in query who paid the highest amounts to Rockbuster.

- Checked the data dictionary, the payment table has the amount paid but customers, like the customer table the payment table had the customer id.
- I added last name, first name, customer id, and payment in the query. Then added the payment table to the join section to extract the amount paid.
- I used WHERE to restrict the data to the top 10 cities.
- Finally, I summarized the data using GROUP BY, ORDER BY and LIMIT to top 5 customers.