Alejandra Castro
28012025

3.7 Table Joins

1. Top 10 countries with the most customers

```
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 country
ORDER BY COUNT(customer_id) DESC
LIMIT 10
```

Data Output | Explain | 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

For this query, we need to find the top ten countries with the most customers. Looking at the ERD, we know that we need to follow this connection of tables: A. customer -> B. address -> C. city -> D. country.

The ERD provides the columns that will connect these tables like so:

- A.customer
    - Customer_id (the item we will be counting)
    - Address_id (A to B connector)
- B.address
    - Address_id (A to B connector)
    - City_id (B to C connector)
- C.city

    - City_id (B to C connector)
    - Country_id (C to D connector)
- D.country
    - Country_id (C to D connector
    - Country (order item, the categories the count will be ordered by)

Additionally, we also know that the result needs to yield a COUNT of customers ORDERED BY country in DESCENDING ORDER for the top 10 items:

- GROUP BY COUNTRY
- ORDER BY COUNT OF customers in DESCENDING ORDER
- LIMIT TO THE 10 TOP RESULTS

2. Find the top 10 cities with the most customers among the top 10 countries

Query Editor | Query History

```sql
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 country IN ('India',
        'China',
        'United States',
        'Japan',
        'Mexico',
        'Brazil',
        'Russian Federation',
        'Philippines',
        'Turkey',
        'Indonesia')
GROUP BY c.city, d.country
ORDER BY COUNT(customer_id) DESC
LIMIT 10
```

Alejandra Castro
28012025

Data Output | Explain | Messages | Notifications

	city character varying (50)	country character varying (50)	number_of_customers bigint
1	Aurora	United States	2
2	Garden Grove	United States	1
3	Salinas	United States	1
4	Araatuba	Brazil	1
5	Talavera	Philippines	1
6	Smolensk	Russian Federation	1
7	Ktahya	Turkey	1
8	Sunnyvale	United States	1
9	Greensboro	United States	1
10	Varanasi (Benares)	India	1

This query required to create a filtered selection of cities based on the top ten countries with the most customers. For this, I only added the column “city” from table C as well as the filter clauses “WHERE” using the top ten countries that were identified in step 1. Additionally, I corrected the table aliases on lines 3 and 18 to match the ones that were given on the SELECT and JOIN clauses.

Looking at the results of the table, it can be inferred that the customers of the top 10 countries are quite spatially scattered. Since these cities belong to the countries with the highest number of customers, and they are the ones that contain the most customers from among these countries, the fact the highest number of customers in a city is 2, followed by all ‘1’ values means that almost none of the customers are concentrated in a single city--not for the top ten and also not for the rest.

This can be proven by running the script again without the cities filter:

Query Editor | Query History

```
1  SELECT c.city,
2         d.country,
3         COUNT(a.customer_id) AS 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 c.city, d.country
9  ORDER BY COUNT(customer_id) DESC
10 LIMIT 10
```

Alejandra Castro
28012025

Data Output | Explain | Messages | Notifications

	city character varying (50)	country character varying (50)	number_of_customers bigint
1	Aurora	United States	2
2	London	United Kingdom	2
3	Greensboro	United States	1
4	Abu Dhabi	United Arab Emirates	1
5	Salala	Oman	1
6	Klerksdorp	South Africa	1
7	Jamalpur	Bangladesh	1
8	Naju	South Korea	1
9	Garden Grove	United States	1
10	Bayugan	Philippines	1

There are only 2 cities in total where there are more than 1 Rockbuster customer, so the rest of the customers are also located at a density of one per city. This indicates the plan to migrate to an online renting service is somewhat compatible. The next step would be to figure out how many customers are serviced by each store and at what location, as well as the number of rentals and revenue per store, in order to figure out rentability.

3. Top 5 customers in the top 10 cities in terms of payment amounts

Query Editor | Query History

```
1  SELECT SUM(A.amount) AS total_payment,
2         B.first_name,
3         B.last_name,
4         D.city,
5         E.country
6  FROM payment A
7  INNER JOIN customer B ON A.customer_id = B.customer_id
8  INNER JOIN address C ON B.address_id = C.address_id
9  INNER JOIN city D ON C.city_id = D.city_id
10 INNER JOIN country E ON D.country_id = E.country_id
11 WHERE city IN ('Aurora',
12                 'Garden Grove',
13                 'Salinas',
14                 'Araatuba',
15                 'Talavera'
16                )
17 GROUP BY A.amount, B.first_name, B.last_name, D.city, E.country
18 ORDER BY total_payment DESC
19 LIMIT 5
```

Alejandra Castro
28012025

Data Output | Explain | Messages | Notifications

	total_payment numeric	first_name character varying (45)	last_name character varying (45)	city character varying (50)	country character varying (50)
1	39.92	Clinton	Buford	Aurora	United States
2	29.94	Rene	Mcalister	Garden Grove	United States
3	29.94	Cassandra	Walters	Salinas	United States
4	26.97	Clinton	Buford	Aurora	United States
5	26.91	Scott	Shelley	Aurora	United States

For this query, the script had to be modified as it had to interact with the payment table, and this has no connection to the tables we had been working with other than just customer. Looking at the ERD, the succession of tables is as follows:

- A.payment
    - Payment_id (primary key)
    - Customer_id (A to B connector)
    - Amount (value to be counted)
- B.customer
    - Customer_id (A to B connector)
    - First_name (item to be identified)
    - Last_name (item to be identified, dependent on first_name)
    - Address_id (B to C connector)
- C.address
    - Address_id (B to C connector)
    - City_id (C to D) connector
- D.city
    - City_id (C to D connector)
    - City (city name, filter)
    - Country_id (D to E connector)
- E.country
    - Country_id (D to E connector)
    - Country (country name)

The query here was to identify the customers from the top 5 cities with the most customers who had paid the most to Rockbuster. For this, I:

1. Created a new SELECT clause to select the amount paid per customer expressed as a sum to represent the total payments, along with the name of the customer, their city, and country.
2. Updated the JOIN clauses to include the new hierarchy detected with the ERD.
3. Used the result of the previous step for the filter clause “WHERE”
4. Grouped them according to the selection clause

Alejandra Castro
28012025

5. Ordered in descending order to identify the top customers.

Looking at the results of the table we can infer there might be some form of data non-uniformity in the payment table because one customer in Aurora appears twice, with two different payment amounts.

To resolve this, the next step would be to create a full join table and inspect this customer, then those in Aurora, and then make random checks on other cities to identify where the problem is and if it only extends to a particular record or several. Although running the script with a limit of 10 already hints at the fact this issue is widespread:

Data Output | Explain | Messages | Notifications

	total_payment numeric	first_name character varying (45)	last_name character varying (45)	city character varying (50)	country character varying (50)
1	39.92	Clinton	Buford	Aurora	United States
2	29.94	Rene	Mcalister	Garden Grove	United States
3	29.94	Cassandra	Walters	Salinas	United States
4	26.97	Clinton	Buford	Aurora	United States
5	26.91	Scott	Shelley	Aurora	United States
6	24.95	Rafael	Abney	Talavera	Philippines
7	23.96	Rafael	Abney	Talavera	Philippines
8	20.97	Cassandra	Walters	Salinas	United States
9	20.97	Rene	Mcalister	Garden Grove	United States
10	20.93	Clinton	Buford	Aurora	United States