Joining tables of Data

1. Write a query to find the top 10 countries for Rockbuster in terms of customer numbers.

I have to associate tables “country” and “customer”. To connect them, I need to join “country” to “city”, then “city” to “address” and finally “address” to “customer”.

Input/query:

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

Output:

"country_name"	**"customer_count"**
"India"	60
"China"	53
"United States"	36
"Japan"	31
"Mexico"	30
"Brazil"	28
"Russian Federation"	28
"Philippines"	20
"Turkey"	15
"Indonesia"	14

I want to see variables “country” and the COUNT of customers, so I applied the respective function. I chose an INNER JOIN because I wanted the query to return **only the rows where matching records existed in both tables.**

In this way, using an INNER JOIN, I only **count customers that are properly linked** across all the relevant tables: country → city → address → customer.

If I would have used a LEFT JOIN, it would have returned **all countries**, even those with **no customers**, and would assign a count of 0 to them.

By using GROUP BY, it’s ensured that customers are count per country; commanding SQL to use aggregate functions like COUNT to compute the numbers

of customers in each group (country); otherwise, it would have just given the total count – all costumers in the entire database – not broken down by country.

By using ORDER BY, it's ensured that the countries with most customers come first, otherwise, the results would appear in random or default order, not by the number of customers.

By using LIMIT 10, it's assured that only the top 10 countries are shown.

2. Next, write a query to identify the top 10 cities that fall within the top 10 countries you identified in step 1. The top 10 cities can be in any of the countries identified—you don't need to create a separate list for each country.

Input/Query: (query in a query aka subquery)

```
SELECT
    city.city AS city_name,
        country.country,
    COUNT(customer.customer_id) AS customer_count
FROM country
INNER JOIN city ON country.country_id = city.country_id
INNER JOIN address ON city.city_id = address.city_id
INNER JOIN customer ON customer.address_id = address.address_id

WHERE country.country_id IN (
  SELECT country.country_id
  FROM country
  JOIN city ON country.country_id = city.country_id
  JOIN address ON city.city_id = address.city_id
  JOIN customer ON customer.address_id = address.address_id
  GROUP BY country.country_id
  ORDER BY COUNT(customer.customer_id) DESC
  LIMIT 10
)

GROUP BY city.city, country.country
ORDER BY customer_count DESC
LIMIT 10;
```

More convenient input:

```
SELECT
    city.city AS city_name,
        country.country,
    COUNT(customer.customer_id) AS customer_count
FROM country
INNER JOIN city ON country.country_id = city.country_id
INNER JOIN address ON city.city_id = address.city_id
INNER JOIN customer ON customer.address_id = address.address_id

WHERE country.country IN ('India', 'China', 'United States', 'Japan', 'Mexico',
'Brazil', 'Russian
Federation', 'Philippines', 'Turkey', 'Indonesia')

GROUP BY city.city, country.country
ORDER BY customer_count DESC
LIMIT 10;
```

Output:

"city_name"	**"country"**	**"customer_count"**
"Aurora"	"United States"	2
"Atlixco"	"Mexico"	1
"Xintai"	"China"	1
"Adoni"	"India"	1
"Dhule (Dhulia)"	"India"	1
"Kurashiki"	"Japan"	1
"Pingxiang"	"China"	1
"Sivas"	"Turkey"	1
"Celaya"	"Mexico"	1
"So Leopoldo"	"Brazil"	1

I want an output that shows city, country and number of costumers. Therefore, using the first query as template, I added the column city to SELECT, kept the INNER JOINs and added a condition, WHERE the cities displayed would be coming from the previous group selection, which means from those 10 top countries. After closing the parenthesis of that condition, I commanded SQL to aggregate the results by city and country using GROUP BY and LIMITed the results for the top 10 cities with most costumers.

3. 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!

Alternative Input:

```
SELECT
    customer.customer_id,
        customer.first_name,
        customer.last_name,
        city.city,
        country.country,
        SUM(payment.amount) AS total_amount_paid
FROM country

INNER JOIN city ON country.country_id = city.country_id
INNER JOIN address ON city.city_id = address.city_id
INNER JOIN customer ON address.address_id = customer.address_id
INNER JOIN payment ON customer.customer_id = payment.customer_id

WHERE city.city IN
('Aurora', 'Atlixco', 'Xintai', 'Adoni', 'Dhule (Dhulia)', 'Kurashiki',
'Pingxiang', 'Sivas', 'Celaya', 'So Leopoldo')
AND country.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,
city.city, country.country
ORDER BY total_amount_paid DESC
LIMIT 5;
```

Output:

"customer_id"	**"first_name"**	**"last_name"**	**"city"**	**"country"**	**"total_amount_paid"**
84	"Sara"	"Perry"	"Atlixco"	"Mexico"	128.70
518	"Gabriel"	"Harder"	"Sivas"	"Turkey"	108.75
587	"Sergio"	"Stanfield"	"Celaya"	"Mexico"	102.76
537	"Clinton"	"Buford"	"Aurora"	"United States"	98.76
367	"Adam"	"Gooch"	"Adoni"	"India"	97.80