3.7 Joining Tables of Data

In this Task you'll get to practice everything you learned in the Exercise. Said simply, you'll need to write a couple of queries combined with joins between the tables address, country, city, customer and payment using their common keys. Create a new text document and call it "Answers 3.7." You'll save your queries, outputs and written answers in this document, as you've done in previous tasks.

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.

```
SELECT D.country,
count (customer_id)
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 (*) desc
limit 10;
```

Data Output | Explain | Messages | Notifi

	country character varying (50)	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.
    a. In order to assess which countries are renting the most movies, we have to first look in the “customer” table. However, the “customer” table only shows us specific customers by name, email, and address, activity, and creation/update data. Since the specific data is within customer and country, we will need to join these tables. Specifically, the address_id designates the exact address of the customer within a “district” (usually state, province, etc) which is in a specific country. The countries are codefied into country_ids which is spelled out in the country table.
    b. Thus, in order to specifically count the number of movie rentals by country, we must combine all four tables to get the bits from each one we need. So we select the country, count by customer_id, we count from the customer table and, after joining with the linking specific data from the tables address, city, and country, we group and order. This combo of queries takes data from all 4 inter-linked tables and combines them so we can count all of the customers from the various regions and specify which purchases were represented by which countries. The result is a list of the number of purchases by country. Then we simply utilize the order function by descending order and limit the count to 10 countries.

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.

```
select C.city,
D.country,
count (customer_id)
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 Federation',
'Philippines',
'Turkey',
'Indonesia')
group by city, country
order by count (*) desc
limit 10;
```

Data Output | Explain | Messages | Notifications

	city character varying (50)	country character varying (50)	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.
    a. This is effectively the same query as above except I added city to the selection and grouping order. The most important distinction is the 'where….in' command so it specified the previously obtained top 10 countries exclusively. That is, London, for example, had 2 rentals and would have showed up in this top 10 list if I hadn't specified the countries listed in the original 10 top 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.

```
select A.customer_id,
```

```
B.first_name,
B.last_name,
D.city,
E.country,
sum(A.amount) as total_amount_paid,
count(A.customer_id) as number_of_rentals
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 city in ('Aurora', 'Atlixco', 'Xintai', 'Adoni', 'Dhule',
 'Kurashiki', 'Pingxiang', 'Sivas', 'Celeya', 'So Leopoldo')
group by A.customer_id, first_name, last_name, city, country
order by sum(amount) desc
limit 5;
```

- Copy-paste your query and its output into your answers document.

Data Output | Explain | 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	number_of_rentals bigint
1	84	Sara	Perry	Atlixco	Mexico	128.70	30
2	518	Gabriel	Harder	Sivas	Turkey	108.75	25
3	537	Clinton	Buford	Aurora	United States	98.76	24
4	367	Adam	Gooch	Adoni	India	97.80	20
5	443	Francisco	Skidmore	So Leopoldo	Brazil	93.79	21

4. Save your “Answers 3.7” document as a PDF and upload it here for your tutor to review.