3.7 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.)
    - 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.

Query | Query History

```
1  Select D.country, count(A.customer_id) as number_of_customers
2  From Customer A
3  Inner join address B on A.address_id =B.address_id
4  Inner join city C on B.city_id = C.city_id
5  Inner join country D on C.country_id = D.country_id
6  Group by D.country
7  Order by count (A.customer_id) Desc
8  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

Total rows: 10 of 10 | Query complete 00:00:00.123

Steps for extracting the top 10 countries in terms of customer numbers

-From the data dictionary I created from the ERD, there are list of tables and name of columns.

-The customer table and country table have information on country and customer ID.

My query is divided into 3 parts : the select, from(join) and summary (using group by, order by and limit 10)

Select

- At the start of the query, I selected the country and counted the number of customer ID's.

From (joining tables)

- Then I joined each table needed to get from the customer table to the country table using inner join. I used two additional tables: city and address table to link the customer and country table.

Summarize data

- Group by country to sum up or summarize for each country, ordered by the number of customers in the country in descending order, and limited to 10 for a top ten countries.

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

```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 D.country IN ('India', 'China', 'United States','Japan','Mexico','Brazil', 'Russian', 'Philippines')
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	Kurashiki	Japan	1
6	Dhule (Dhulia)	India	1
7	Pingxiang	China	1
8	Ozamis	Philippines	1
9	Nezahualcyotl	Mexico	1
10	So Leopoldo	Brazil	1

Steps to write the query

- I checked the data dictionary from ERD to know which table is required for this query.
- Similar to the first query but I added city variable in the select section of the query and then used a WHERE and IN syntax to restrict the list of countries to the top 10 in the output in the first query.
  Looking at the output there is only one city with more than one customer, the Aurora city in United states had two customers. The remaining 9 cities had 1 customer.

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

```
1  Select A.customer_id,
2  A.first_name,
3  A.last_name,
4  D.city,
5  E.country,
6  SUM(B.amount) as total_amount_paid
7  From Customer A
8  Inner join payment B on A.customer_id = B.customer_id
9  Inner join address c on A.address_id = C.address_id
10 Inner join city D on C.city_id = D.city_id
11 Inner join country E on D.country_id = E.country_id
12 where D.city IN ('Aurora', 'Atlixco','Xintai', 'Adoni', 'Dhule (Dhulia)',
13                  'Kurashiki','Pingxang', 'Sivas', 'Celaya', 'So Leopoldo')
14 Group by A.customer_id, A.first_name, A.last_name, D.city, E.Country
15 Order by  total_amount_paid Desc
16 Limit 5
```

	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	84	Sara	Perry	Atlixco	Mexico	128.7
2	518	Gabriel	Harder	Sivas	Turkey	108.75
3	587	Sergio	Stanfield	Celaya	Mexico	102.76
4	537	Clinton	Buford	Aurora	United States	98.76
5	367	Adam	Gooch	Adoni	India	97.8

Steps to write the query

In the third query we want to extract the top 5 customers in the top 10 cities in query 2 who paid the highest amounts to rockuster.

- I checked the data dictionary, and the payment table has the amount paid by customers. Similar to the customer table the payment table had the customer id variable.
- I tweaked the second query I included last name, first name, customer id, and payment in the select section. Then added the payment table to the join section of the query to extract the amount paid.
- I used where clause to restrict the data to the top 10 cities in query 2.
- Then, I summarize the data using group by, order by and limit to 5 top customers.