Task 3.7 – Joining Tables of Data

Rockbuster's management team would like to know the top 10 countries where customers are based.

1.

```
SELECT D.country, COUNT(A.customer_id) AS num_of_customers
FROM customer A
JOIN address B ON A.address_id = B.address_id
JOIN city C ON B.city_id = C.city_id
JOIN country D ON C.country_id = D.country_id
GROUP BY country
ORDER BY num_of_customers DESC
LIMIT 10
```

country character varying (50)	num_of_customers bigint
India	60
China	53
United States	36
Japan	31
Mexico	30
Brazil	28
Russian Federation	28
Philippines	20
Turkey	15
Indonesia	14

Since the question is asking to find the top 10 **countries** in terms of **customer numbers**, it is looking for two columns. The first step was to refer to the ERD that was created in previous tasks. I can see that the **customer** table is linked to the **address** table, the **address** table is linked to the **city** table, and the **city** table is linked to the **country** table.

Because I want limited information from just two tables, I know to use INNER JOIN. With the tip provided, since I am selecting the country column, I also have to GROUP BY that same column. It's asking for the top 10 countries so I used ORDER BY DESC to get the highest number first. Last but not least, it asked for 10 countries, so the LIMIT 10 command was used.

2.

```
SELECT C.city, D.country, COUNT(A.customer_id) AS num_of_customers
FROM customer A
JOIN address B ON A.address_id = B.address_id
JOIN city C ON B.city_id = C.city_id
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 city, country
ORDER BY num_of_customers DESC
LIMIT 10
```

city character varying (50)	country character varying (50)	num_of_customers bigint
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

Because it's using the query from the previous question, I know that I have to modify it to find the top 10 cities, with city being the keyword.

Using the same query, I added C.city so I can add another column displaying the cities. The joins remained the same because based on the ERD, these are how the tables are connected. But since I know what the top 10 countries are, I added a WHERE clause to be specific and filter out the other countries. Since I am now selecting an additional column (city), I added it to GROUP BY clause otherwise I'd get an error.

3.

```
SELECT A.customer_id,
       first_name,
       last_name,
       E.country,
       D.city,
       SUM(B.amount) AS total_amount_paid
FROM customer A
JOIN payment B ON A.customer_id = B.customer_id
JOIN address C ON A.address_id = C.address_id
JOIN city D ON A.city_id = D.city_id
JOIN country E ON A.country_id = E.country_id
WHERE city IN ('Aurora', 'Atlixco', 'Xintai', 'Adoni',
'Dhule (Dhulia)', 'Kurashiki', 'Pingxiang', 'Sivas', 'Celaya',
'So Leopoldo')
GROUP BY A.customer_id, first_name, last_name, country, city
ORDER BY total_amount_paid DESC
LIMIT 5
```

customer_id integer	first_name character varying (45)	last_name character varying (45)	country character varying (50)	city character varying (50)	total_amount_paid numeric
84	Sara	Perry	Mexico	Atlixco	128.70
518	Gabriel	Harder	Turkey	Sivas	108.75
587	Sergio	Stanfield	Mexico	Celaya	102.76
537	Clinton	Buford	United States	Aurora	98.76
367	Adam	Gooch	India	Adoni	97.80

Based on the tip, it mentions that the output should include specific columns. From that, I know what columns I must SELECT and the order I want to present it in. I can see that there's an "amount" column in the "payment" table, so I use the aggregate function (SUM) to total the amount and give it an alias.

I then join the tables, by first connecting "customer" and "payment" as they are both connected by the customer_id. From there, I connected the customer table to the address, city, and country table as I did the previous two questions. Since I know the top 10 cities, I used the WHERE clause to filter out the other cities. Because I selected very specific columns, I must also use GROUP BY. And finally, using ORDER BY to rank them from highest to lowest and limiting the results to 5.