Databases & SQL for Analysts

Task 3.7: Joining Tables of Data

Ngawang Dhundup

LINK NAME:

Directions:

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(A.customer_id) AS customer_count
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 D.country
ORDER BY Count(A.customer_id) desc
LIMIT 10
```

	country character varying (5	customer_count
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

Save results to file
F8

- 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.
  - I pulled up the ERD and wrote down the tables and columns needed for the query. Wrote down a flow chart so I could understand how many tables and in what sequence I needed to join them. Once I had all the keys, I decided to use inner join since I was only interested in the matching values of customers and countries. Joined the 4 tables according to each key, then grouped by country and used count function for customer id. Then ordered by descending order, highest count to lowest, and limited the output to just top ten.

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(A.customer_id) AS customer_count
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')
GROUP BY C.city, D.country
ORDER BY Count(A.customer_id) DESC
LIMIT 10;
```

    - Write a short explanation of how you approached this query and why.
        - I just added the city from table C (city) to the select and group by statement from the previous step. I also added the WHERE clause to limit the results to the top ten countries I received earlier.

	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	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

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

```
SELECT
A.customer_id,
A.first_name AS customer_first_name,
A.last_name AS customer_last_name,
A.email AS customer_email,
C.city,
D.country,
SUM(E.amount) AS total_amount_paid
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
INNER JOIN payment E ON A.customer_id = E.customer_id
WHERE C.city IN ('Aurora',
'Atlixco',
'Xintai',
'Adoni',
'Kurashiki',
'Dhule (Dhulia)',
'Pingxiang',
'Ozamis',
'Nezahualcyotl',
'So Leopoldo')
GROUP BY
A.customer_id,
A.first_name,
A.last_name,
A.email,
C.city, D.country
```

```
ORDER BY total_amount_paid desc
LIMIT 5;
```

Data output | Messages | Notifications

	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	128.70
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	97.80

Total rows: 5 of 5 Query complete 00:00:00.056

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