Omid Soltani
Data Analytics Immersion – Task 3.7

Step 1

Query | Query History

```
1  SELECT
2      country.country_id,
3      country,
4      COUNT(DISTINCT customer_id) AS customer_count
5  FROM
6      customer
7  JOIN
8      address ON customer.address_id = address.address_id
9  JOIN
10     city ON address.city_id = city.city_id
11 JOIN
12     country ON city.country_id = country.country_id
13 GROUP BY
14     country.country_id, country
15 ORDER BY
16     customer_count DESC
17 LIMIT
18     10;
```

Data Output | Messages | Notifications

	country_id [PK] integer	country character varying (50)	customer_count bigint
1	44	India	60
2	23	China	53
3	103	United States	36
4	50	Japan	31
5	60	Mexico	30
6	15	Brazil	28
7	80	Russian Federation	28
8	75	Philippines	20
9	97	Turkey	15
10	45	Indonesia	14

Total rows: 10 of 10 | Query complete 00:00:00.045

My first step was to identify the necessary tables. I focused on connecting the 'customer' table with the 'address', 'city', and 'country' tables using the corresponding foreign keys.

I used the JOIN statements to link these tables based on their primary and foreign key relationships.

The COUNT function helped me calculate the distinct number of customers in each country.

Additionally, the GROUP BY clause was essential to aggregate the data by country, and the ORDER BY clause arranged the results in descending order by customer count.

Finally, the LIMIT clause allowed me to extract only the top 10 countries.

Step 2

Query | Query History

```
1  SELECT
2      city.city_id,
3      city.city,
4      COUNT(DISTINCT customer.customer_id) AS customer_count
5  FROM
6      customer
7  JOIN
8      address ON customer.address_id = address.address_id
9  JOIN
10     city ON address.city_id = city.city_id
11 JOIN
12     country ON city.country_id = country.country_id
13 GROUP BY
14     city.city_id, city.city
15 ORDER BY
16     customer_count DESC
17 LIMIT
18     10;
```

Data Output | Messages | Notifications

	city_id [PK] integer	city character varying (50)	customer_count bigint
1	42	Aurora	2
2	312	London	2
3	3	Abu Dhabi	1
4	5	Adana	1
5	6	Addis Abeba	1
6	7	Aden	1
7	9	Ahmadnagar	1
8	8	Adoni	1
9	10	Akishima	1
10	4	Acua	1

Total rows: 10 of 10 | Query complete 00:00:00.061

I began by selecting the relevant tables ('customer', 'address', 'city', and 'country') and establishing their relationships and then JOIN the tables based on their common keys.

Next, I implemented the GROUP BY clause to group the results by city, ensuring that each city appears only once in the output.

The COUNT(DISTINCT customer.customer_id) allowed me to calculate the number of distinct customers in each city, which I sorted in descending order to identify the top cities.

Finally, I applied LIMIT 10 so that only the top 10 cities would be included in the output.

Omid Soltani
Data Analytics Immersion – Task 3.7

Step 3

Query | Query History

```sql
SELECT
    customer.customer_id,
    customer.first_name,
    customer.last_name,
    country.country,
    city.city,
    SUM(payment.amount) AS total_amount_paid
FROM
    customer
JOIN
    payment ON customer.customer_id = payment.customer_id
JOIN
    address ON customer.address_id = address.address_id
JOIN
    city ON address.city_id = city.city_id
JOIN
    country ON city.country_id = country.country_id
WHERE
    city.city IN ('Aurora', 'London', 'Abu Dhabi', 'Adana', 'Addis Abeba')
GROUP BY
    customer.customer_id, customer.first_name, customer.last_name, country.country, city.city
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)	country character varying (50)	city character varying (50)	total_amount_paid numeric
1	512	Cecil	Vines	United Kingdom	London	115.74
2	452	Tom	Milner	United Arab Emirates	Abu Dhabi	99.71
3	537	Clinton	Buford	United States	Aurora	98.76
4	392	Sean	Douglass	Ethiopia	Addis Abeba	91.77
5	327	Larry	Thrasher	Turkey	Adana	74.81

Total rows: 5 of 5 | Query complete 00:00:00.052