

# User Experience Percentage

## Question

Given a table called `user_experiences`, write a query to determine the percentage of users that held the title of “Data Analyst” immediately before holding the title “Data Scientist”.

Immediate is defined as the user holding no other titles between the “Data Analyst” and “Data Scientist” roles.

### Input:

`user_experiences` table

Column	Type
<code>id</code>	INTEGER
<code>position_name</code>	VARCHAR
<code>start_date</code>	DATETIME
<code>end_date</code>	DATETIME
<code>user_id</code>	INTEGER

### Output:

Column	Type
<code>percentage</code>	FLOAT

## Solution

```
WITH previous_role AS (  
    SELECT  
        user_id, position_name,  
        LAG (position_name) OVER(partition by user_id) AS prev_position  
    FROM user_experiences  
),  
immediate_subset AS (  
    SELECT  
        count(distinct user_id) AS nr_occurrences  
    FROM previous_role  
    WHERE position_name = "Data Scientist"  
        AND prev_position = "Data Analyst"  
)  
  
SELECT  
    nr_occurrences / (SELECT  
                        COUNT(DISTINCT user_id)  
                        FROM user_experiences) AS percentage  
FROM immediate_subset
```

## Solution Approach

1. Create a new column using the LAG window function. The LAG function returns the value of the position\_name from the row that precedes the current row. The PARTITION BY clause divides the rows in the result-set into partitions according to each user\_id .
2. We define a subset of the users who satisfied the condition : the specific rows that held the "Data Analyst" title immediately before "Data Scientist" and then count the distinct number of occurrences of these rows. This is aliased as nr\_occurrences
3. In the main query we compute the percentage as the nr\_occurrences/ the total number of all distinct user\_id's. We give this the name percentage.

```
WITH previous AS (  
    SELECT *,  
    LAG(position_name, 1) OVER (PARTITION BY user_id ORDER BY start_date) previous  
    FROM user_experiences  
)  
  
SELECT COUNT(DISTINCT user_id) / (SELECT COUNT(DISTINCT user_id) FROM previous) percentage  
FROM previous  
WHERE position_name = "Data Scientist" AND previous = "Data Analyst"
```