



Data Science Skills (LinkedIn)

Given a table of candidates and their skills, you're tasked with finding the candidates best suited for an open Data Science job. You want to find candidates who are proficient in Python, Tableau, and PostgreSQL.

Write a query to list the candidates who possess all of the required skills for the job. Sort the output by candidate ID in ascending order.

Assumption:

- There are no duplicates in the `candidates` table.

`candidates` Table:

Column Name	Type
candidate_id	integer
skill	varchar

`candidates` Example Input:

candidate_id	skill
123	Python
123	Tableau
123	PostgreSQL
234	R
234	PowerBI
234	SQL Server

345	Python
345	Tableau

Step 1: Identify the problem of the case

Find candidates who are proficient in Python, Tableau, and PostgreSQL.

The Output includes id but qualified candidate, and sorted in ascending id order.

Column Name	Type
candidate_id	int

Step 2 : Analyze and solve problems

If you use the `COUNT()` function with `FILTER`, you can only get one condition in a cell

We use the `WHERE` function to filter the records that contain 'Python', 'Tableau', 'PostgreSQL', the type of features beyond the 3 skills above.

```
SELECT *
FROM candidates
WHERE skill IN ('Python', 'Tableau', 'PostgreSQL')
```

Next, use Window Function with `COUNT()` to count the number of skills that the candidate has in the 3 skills above.

```
SELECT *
, COUNT(candidate_id) OVER(PARTITION BY candidate_id ORDER BY skill) num_skill
FROM candidates
WHERE skill IN ('Python', 'Tableau', 'PostgreSQL')
```

Finally from this view we filter the id of the candidates with the number of skills greater than 3.

```
SELECT candidate_id
FROM(
SELECT *
, COUNT(candidate_id) OVER(PARTITION BY candidate_id ORDER BY skill) num_skill
FROM candidates
WHERE skill IN ('Python', 'Tableau', 'PostgreSQL'))
```

```
) sub  
WHERE num_skill = 3
```

We have the result:

candidate_id
123
147