

(Linkedin) Matching Skills - Easy Solution

Source - <https://datalemur.com/questions/matching-skills>

Running notes

Tables we have -

candidate | skill

Question

1. find candidates proficient in Python, Tableau, and PostgreSQL (ALL 3 SKILLS)
2. Sort the output by candidate ID in ascending order

Solution

```
SELECT *  
FROM candidates;
```

This has given me an overall view of the database, so i know how the dataset looks like

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

This is now showing me all the candidates that have any one of these skills



WHERE clause outputs the rows where a certain condition is met



IN clause

- used to filter categorical data
- instead of a more tedious approach of multiple ORs we can directly use an IN clause
- so here it is checking if anything from the 'skill' column is matching to either - Python, Tableau or PostgreSQL and if any of the value is matched that row is shown in the output

Now the question further says that I need to find those candidate_id's who have all the 3 skills

```
SELECT candidate_id, COUNT(skill) as number_of_skills_out_of_3_1
FROM candidates
WHERE skill IN ('Python', 'Tableau', 'PostgreSQL')
GROUP BY candidate_id;
```

So in order to group by the candidates and count the number of skills they have from the 3 required skills



GROUP BY clause tells the database to separate the data in different groups so we can perform aggregation on these groups separately (like how I need to perform an aggregation on the number of skills to see if they have all the 3 required skills)



COUNT counts how many rows we have in a particular column

Output



candidate_id	number_of_skills_out_of_3_required_skills
123	3
345	2

so this is how my output looks like so now I know candidate 123 has all the 3 required skills but candidate 345 has only 2 required skills

so my output should only show candidate_id 123

```
SELECT candidate_id, COUNT(skill) as number_of_skills_out_of_3_  
FROM candidates  
WHERE skill IN ('Python', 'Tableau', 'PostgreSQL')  
GROUP BY candidate_id  
WHERE COUNT(skill) = 3;
```

So a general thought would lead us to using a WHERE clause (WHERE clause outputs only a certain rows where the condition is satisfied)

This code will give an error because *aggregate functions are not allowed in WHERE clause*



WHERE clause is used to filter data before aggregation



To filter data based on an aggregate function result, we must use the **HAVING clause.**

So, drumrolls for the final query

```
SELECT candidate_id  
FROM candidates
```

```
WHERE skill IN ('Python', 'Tableau', 'PostgreSQL')  
GROUP BY candidate_id  
HAVING COUNT(skill) = 3  
ORDER BY candidate_id;
```

| Written By,

| Harshee Pitroda