DataLemur

‹ Back to questions

Duplicate Job Listings

Description	Solution	Discussion	Submissions

Easy   Linkedin

Share on Twitter   **Share on LinkedIn**

This is the same question as problem #8 in the SQL Chapter of Ace the Data Science Interview!

Assume you are given the table below that shows job postings for all companies on the LinkedIn platform. Write a query to get the number of companies that have posted duplicate job listings.

Clarification:

- Duplicate job listings refer to two jobs at the same company with the same title and description.

`job_listings` **Table:**

Column Name	Type
job_id	integer
company_id	integer

```
1  SELECT count(DISTINCT company_id) FROM
2  (SELECT company_id,title,description,COUNT(job_id)
3  as job_count
4  FROM job_listings
5  GROUP BY company_id,title,description)
6  AS duplicate_jobs WHERE job_count>1;
```

description	string

`job_listings` Example Input:

job_id	company_id	title	description
248	827	Business Analyst	Business analyst evaluates past and current business data with the primary goal of improving decision-making processes within organizations.
149	845	Business Analyst	Business analyst evaluates past and current business data with the primary goal of improving decision-making processes within organizations.
945	345	Data Analyst	Data analyst reviews data to identify key insights into a business's customers and ways the data can be used to solve problems.
164	345	Data Analyst	Data analyst reviews data to identify key insights into a business's customers and ways the data can be used to solve problems.

			Data engineer works in a
172	244	Data Engineer	variety of settings to build systems that collect, manage, and convert raw data into usable information for data scientists and business analysts to interpret.

Example Output:

co_w_duplicate_jobs
1

Explanation

Because job IDs 945 and 164 are at the same company (345), and the jobs have the same title and description, there is exactly one company with a duplicate job.

The dataset you are querying against may have different input & output - **this is just an example**!

Gimme a Hint

PostgreSQL 14

Run Code   Submit

Output   ⌄

count
3