

Duplicate Job Listings [Linkedin SQL Interview Question]

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

Assume you're given a table containing job postings from various companies on the LinkedIn platform. Write a query to retrieve the count of companies that have posted duplicate job listings.

Definition:

- Duplicate job listings are defined as two job listings within the same company that share identical titles and descriptions.

`job_listings` Table:

Column Name	Type
job_id	integer
company_id	integer
title	string
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.

job_id	company_id	title	description
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.
172	244	Data Engineer	Data engineer works in a 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:

duplicate_companies
1

Explanation:

There is one company ID 345 that posted duplicate job listings. The duplicate listings, IDs 945 and 164 have identical titles and descriptions.

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