

DataLemur - Data Science Skills

LinkedIn SQL Interview Question

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

Example Output:	
candidate_id	
123	

Solution:

```
WITH total_skill AS
(select candidate_id , count(skill) as total_skill from candidates
where skill in ('Python', 'Tableau', 'PostgreSQL')
group by candidate_id
)
```

```
select candidate_id from total_skill
where total_skill >=3
```

DataLemur - Histogram of Tweets

Twitter SQL Interview Question

Assume you're given a table Twitter tweet data, write a query to obtain a histogram of tweets posted per user in 2022. Output the tweet count per user as the bucket and the number of Twitter users who fall into that bucket.

In other words, group the users by the number of tweets they posted in 2022 and count the number of users in each group.

tweets Table:

Column Name	Type
tweet_id	integer
user_id	integer
msg	string
tweet_date	timestamp

tweets Example Input:

tweet_id	user_id	msg	tweet_date
214252	111	Am considering taking Tesla private at \$420. Funding secured.	12/30/2021 00:00:00
739252	111	Despite the constant negative press covfefe	01/01/2022 00:00:00
846402	111	Following @NickSinghTech on Twitter changed my life!	02/14/2022 00:00:00
241425	254	If the salary is so competitive why won't you tell me what it is?	03/01/2022 00:00:00
231574	148	I no longer have a manager. I can't be managed	03/23/2022 00:00:00

Example Output:

tweet_bucket	users_num
1	2
2	1

Solution:

```
WITH total_tweets AS (  
  SELECT  
    user_id,  
    COUNT(tweet_id) AS tweet_count_per_user  
  FROM tweets  
  WHERE tweet_date BETWEEN '2022-01-01'  
    AND '2022-12-31'  
  GROUP BY user_id)
```

```
SELECT  
  tweet_count_per_user AS tweet_bucket,  
  COUNT(user_id) AS users_num  
FROM total_tweets  
GROUP BY tweet_count_per_user;
```

Datalemur - Page With No Likes

Facebook SQL Interview Question

Assume you're given two tables containing data about Facebook Pages and their respective likes (as in "Like a Facebook Page").

Write a query to return the IDs of the Facebook pages that have zero likes. The output should be sorted in ascending order based on the page IDs.

pages Table:

Column Name	Type
page_id	integer
page_name	varchar

pages Example Input:

page_id	page_name
20001	SQL Solutions
20045	Brain Exercises
20701	Tips for Data Analysts

page_likes Table:

Column Name	Type
user_id	integer
page_id	integer
liked_date	datetime

page_likes Example Input:

user_id	page_id	liked_date
111	20001	04/08/2022 00:00:00
121	20045	03/12/2022 00:00:00
156	20001	07/25/2022 00:00:00

Solution:

```
SELECT page_id FROM pages
WHERE page_id not in
(select DISTINCT page_id
from page_likes)
```

Datalemur Unfinished Parts

Tesla SQL Interview Question

Tesla is investigating production bottlenecks and they need your help to extract the relevant data. Write a query to determine which parts have begun the assembly process but are not yet finished.

Assumptions:

- `parts_assembly` table contains all parts currently in production, each at varying stages of the assembly process.
- An unfinished part is one that lacks a `finish_date`.

This question is straightforward, so let's approach it with simplicity in both thinking and solution.

Effective April 11th 2023, the problem statement and assumptions were updated to enhance clarity.

`parts_assembly` Table

Column Name	Type
part	string
finish_date	datetime
assembly_step	integer

`parts_assembly` Example Input

part	finish_date	assembly_step
battery	01/22/2022 00:00:00	1
battery	02/22/2022 00:00:00	2
battery	03/22/2022 00:00:00	3
bumper	01/22/2022 00:00:00	1
bumper	02/22/2022 00:00:00	2
bumper		3
bumper		4

Solution:

```
SELECT part, assembly_step FROM parts_assembly
where finish_date is null
```

Datalemur: Laptop vs. Mobile Viewership

New York Times SQL Interview Question

Assume you're given the table on user viewership categorised by device type where the three types are laptop, tablet, and phone.

Write a query that calculates the total viewership for laptops and mobile devices where mobile is defined as the sum of tablet and phone viewership. Output the total viewership for laptops as `laptop_reviews` and the total viewership for mobile devices as `mobile_views`.

viewership Table

Column Name	Type
user_id	integer
device_type	string ('laptop', 'tablet', 'phone')
view_time	timestamp

viewership Example Input

user_id	device_type	view_time
123	tablet	01/02/2022 00:00:00
125	laptop	01/07/2022 00:00:00
128	laptop	02/09/2022 00:00:00
129	phone	02/09/2022 00:00:00
145	tablet	02/24/2022 00:00:00

Solution:

```
select
COUNT(CASE
  WHEN device_type = 'laptop' THEN 1
  ELSE NULL
END) AS laptop_views,
COUNT(CASE
  WHEN device_type = 'tablet' or device_type = 'phone' THEN 1
  ELSE NULL
END) AS mobile_views
from viewership
```