

Histogram of Tweets [Twitter SQL Interview Question]

This is the same question as problem #6 in the SQL Chapter of [Ace the Data Science Interview!](#)

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

tweet_id	user_id	msg	tweet_date
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

Explanation:

Based on the example output, there are two users who posted only one tweet in 2022, and one user who posted two tweets in 2022. The query groups the users by the number of tweets they posted and displays the number of users in each group.

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