DataLemur Medium

Tweets' Rolling Averages [Twitter SQL Interview Question]

Important Assumptions:

- Rows in this table are *consecutive* and ordered by date.
- Each row represents a different day
- A day that does not correspond to a row in this table is not counted. The most recent day is the next row above the current row.

Note: Rolling average is a metric that helps us analyze data points by creating a series of averages based on different subsets of a dataset. It is also known as a moving average, running average, moving mean, or rolling mean.

`tweets` Table:

Column Name	Type
tweet_id	integer
user_id	integer
tweet_date	timestamp

`tweets` Example Input:

tweet_id	user_id	tweet_date
214252	111	06/01/2022 12:00:00
739252	111	06/01/2022 12:00:00
846402	111	06/02/2022 12:00:00
241425	254	06/02/2022 12:00:00
137374	111	06/04/2022 12:00:00

Example Output:

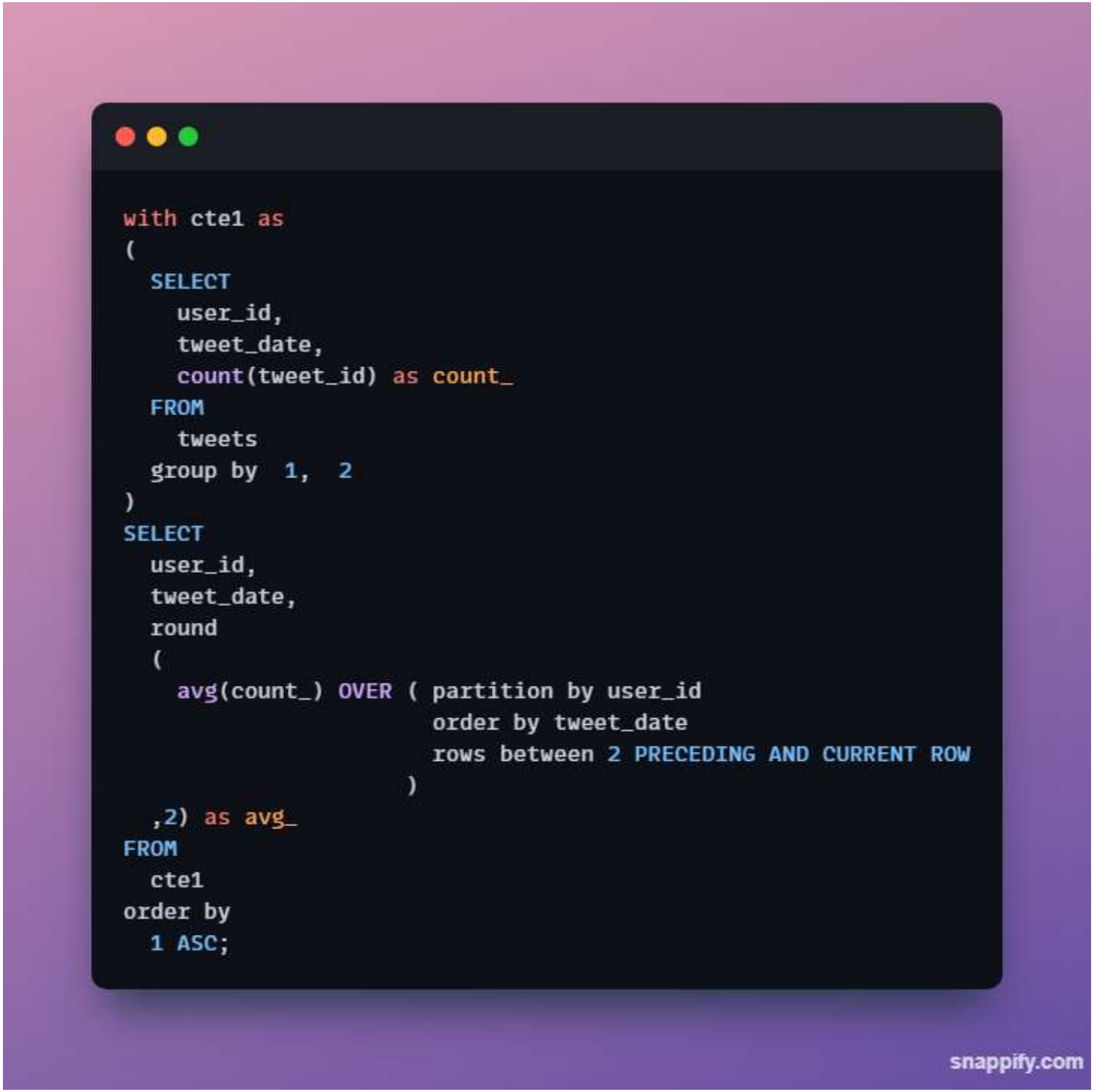
user_id	tweet_date	rolling_avg_3days
111	06/01/2022 12:00:00	2.00
111	06/02/2022 12:00:00	1.50
111	06/04/2022 12:00:00	1.33
254	06/02/2022 12:00:00	1.00

Explanation

User 111 made 2 tweets on 06/01/2022, and 1 tweet the next day. By 06/02/2022, the user had made in total 3 tweets over the course of 2 days; thus, the rolling average is 3/2=1.5. By 06/04/2022, there are 4 tweets that were made during 3 days: 4/3 = 1.33 rolling average.

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

SOLUTION QUERY--->


with cte1 as
(
SELECT
user_id,
tweet_date,
count(tweet_id) as count_
FROM
tweets
group by 1, 2
)
SELECT
user_id,
tweet_date,
round
(
avg(count_) OVER (partition by user_id
order by tweet_date
rows between 2 PRECEDING AND CURRENT ROW
)
,2) as avg_
FROM
cte1
order by
1 ASC;
snappify.com