



Average Post Hiatus p1 (Facebook)

QUESTION

Given a table of Facebook posts, for each user who posted at least twice in 2021, write a query to find the number of days between each user's first post of the year and last post of the year in the year 2021. Output the user and number of the days between each user's first and last post.

p.s. If you've read the [Ace the Data Science Interview](#) and liked it, consider writing us a review?

posts Table:

Column Name	Type
user_id	integer
post_id	integer
post_date	timestamp
post_content	text

posts Example Input:

user_id	post_id	post_date	post_content
151652	599415	07/10/2021 12:00:00	Need a hug
661093	624356	07/29/2021 13:00:00	Bed. Class 8-12. Work 12-3. Gym 3-5 or 6. Then class 6-10. Another day that's gonna fly by. I miss my girlfriend
004239	784254	07/04/2021 11:00:00	Happy 4th of July!
661093	442560	07/08/2021 14:00:00	Just going to cry myself to sleep after watching Marley and Me.
151652	111766	07/12/2021 19:00:00	I'm so done with covid - need travelling ASAP!

Step 1: Identify the problem of the case

The case asks to find the number of days between the user's first post of the year and the last post of 2021.

Output will have:

Column Name	Type
user_id	integer
day_between	integer

Step 2 : Analyze and solve problems

To find the number of days between each user's first post of the year and last post of the year in the year 2021, we need to find the first and last day of the year, we will use the `MAX()` and `MIN()` and use `WHERE` to filter the data in 2021.

Then number of days between each user's first post of the year and last post of the year = the last day minus the first day, use the `EXTRACT()` function to get the date result.

```
SELECT user_id
      ,EXTRACT(day FROM (MAX(post_date) - MIN(post_date))) days_between
   FROM posts
 WHERE EXTRACT(year FROM post_date) = 2021
 GROUP BY user_id
```

Nearly done! the result :

```
SELECT user_id, days_between
FROM (
  SELECT user_id
    ,EXTRACT(day FROM (MAX(post_date) - MIN(post_date))) days_between
   FROM posts
 WHERE EXTRACT(year FROM post_date) = 2021
 GROUP BY user_id
) SUB
WHERE days_between > 0
ORDER BY days_between DESC
```

And the output table:

user_id	days_between
151652	307
661093	206