9_Average Post Hiatus (Part 1) Easy

Source - nicksinghtech **Average Post Hiatus | Facebook SQL Interview Question**

Running Notes

- 1 table - posts

Query

- find the users who posted atleast twice in 2021
- find the number of days between the first post and last post

Solution

- people who posted twice their user id will come twice or more than twice so the below query does that

```
SELECT user_id FROM posts GROUP BY user_id HAVING count(user_id)>=2;
```

- and then to find the number of days between the first post and last post I am finding the difference between the max post date and min post date and converting it to year

```
SELECT user_id, DATE_PART('days',MAX(post_date) - MIN(post_date)) AS
days_between FROM posts GROUP BY user_id HAVING count(user_id)>=2;
```

I submitted this query and it wasnt accepted, aaaaaaaaaaaaaaaaa

well well i missed a small condition on re-reading the question i noticed what i missed was the year 2021

```
SELECT user_id, DATE_PART('days',MAX(post_date) - MIN(post_date)) AS
days_between FROM posts WHERE DATE_PART('year',post_date) = '2021' GROUP
BY user_id HAVING count(user_id)>=2;
```

- i first added the DATE_PART('year',post_date) = '2021' in the HAVING clause
  The issue with placing the `DATE_PART('year', post_date) = '2021'` condition in the `HAVING` clause is that the `HAVING` clause is used to filter groups **after** the aggregation (such as `COUNT`, `MAX`, `MIN`), whereas the `WHERE` clause filters rows **before** any aggregation happens.

> 💡 `WHERE` clause is used to filter rows **before** any aggregation
> It works on individual rows and allows you to set conditions based on the raw data.
>
> We use the `WHERE` clause to filter posts from a specific year or posts from a particular user before grouping them.

> 💡 The `HAVING` clause is used to filter data on the results of aggregate functions.
>
> We use `HAVING` to set conditions based on aggregated values, like filtering for groups where the total number of posts (`COUNT(user_id)`) is greater than a threshold

so the final query that got accepted

```
SELECT user_id, DATE_PART('days',MAX(post_date) - MIN(post_date)) AS
days_between FROM posts WHERE DATE_PART('year',post_date) = '2021' GROUP
BY user_id HAVING count(user_id)>=2;
```

Written By,

Harshee Pitroda