

Problem:

A buyer is ***new*** on the day of their first-ever purchase.

A buyer is ***reactivated*** if they purchase after ≥ 30 days of inactivity.

Write a query to get the number of ****new**** and ****reactivated**** buyers on each day between Nov 1–7 2025.

****Table:**** trans_table(seller_id, buyer_id, trans_dt, category, sale_value)

Expected Output:

day	new_buyers	reactivated_buyers
2025-11-01	220	40
2025-11-02	260	35
2025-11-03	230	42
2025-11-04	210	50
2025-11-05	200	45
2025-11-06	195	38
2025-11-07	240	52

Learning: →

Re-activation - is evaluated per buyer per day, not per buyer over entire history.

- 1) Use Lag(prev_purchase_date)
- 2) Not date_diff(end_date, start_date) = 1

Critical failing scenario:

buyer_id	trans_dt
404	2025-10-01
404	2025-11-01
404	2025-11-04

min(dt) as start_dt - max(dt) as end_dt

What should happen:

- 1) Nov 1:
- Previous = Oct 1 -> gap = 31 -> reactivated.
- 2) Nov 4:
- Previous = Nov 1 -> gap = 3 -> Not reactivated.

Your query(previous one):

- start_dt = 2025-10-01
- end_dt = 2025-11-04

- datediff = 34 -> reactivated for both days (which is wrong)

Your method can't distinguish the **daily reactivation status**

Soln:

```
with ordered AS(
select
buyer_id,
DATE(trans_dt) as day,
LAG(date(trans_dt)) OVER(PARTITION BY buyer_id ORDER BY DATE(trans_dt)) AS
prev_purchase
from trans_table
),
classified AS(
select
day,
buyer_id,
case
when prev_purchase is null then 'new'
when DATE_DIFF(day, prev_purchase, day) >= 30 THEN 'reactivated'
else 'other'
end as buyer_type
from ordered
)
select
day,
COUNT(DISTINCT CASE WHEN buyer_type = 'new' THEN buyer_id END) AS new_buyers,
COUNT(DISTINCT CASE WHEN buyer_type = 'reactivated' THEN buyer_id END) AS
reactivated_buyers
from classified
where day BETWEEN '2025-11-01' and '2025-11-07'
group by day
order by day
```

How LAG solves it:

buyer	trans_dt	prev_purchase	gap
----	-----	-----	--
404	2025-10-01	NULL	-
404	2025-11-01	2025-10-01	31

