DECLARE date\_latest\_tweet, date\_latest\_follows DATE;

SET date\_latest\_tweet = (

SELECT PARSE\_DATE('%Y%m%d', SUBSTRING(MAX(partition\_id), 1, 8)) AS partition\_id

FROM `twttr-bq-tweetsource-pub-prod.user.INFORMATION\_SCHEMA.PARTITIONS`

WHERE partition\_id IS NOT NULL AND partition\_id != '\_\_NULL\_\_' AND table\_name="public\_tweets");

SET date\_latest\_follows = (

SELECT PARSE\_DATE('%Y%m%d', MAX(partition\_id)) AS partition\_id

FROM `twttr-recos-ml-prod.user\_events.INFORMATION\_SCHEMA.PARTITIONS`

WHERE partition\_id IS NOT NULL AND partition\_id != '\_\_NULL\_\_' AND table\_name="valid\_user\_follows");

-- tweet count candidate features

CREATE OR REPLACE TABLE `twttr-recos-ml-prod.realgraph.tweeting\_follows`

PARTITION BY ds

AS

WITH tweet\_count AS (

SELECT userId, COUNT(userId) AS num\_tweets

FROM `twttr-bq-tweetsource-pub-prod.user.public\_tweets`

WHERE DATE(ts) BETWEEN DATE\_SUB(date\_latest\_tweet, INTERVAL 3 DAY) AND date\_latest\_tweet

GROUP BY 1

), all\_follows AS (

SELECT F.sourceId AS source\_id, F.destinationId AS destination\_id, COALESCE(T.num\_tweets,0) AS num\_tweets,

ROW\_NUMBER() OVER (PARTITION BY F.sourceId ORDER BY T.num\_tweets DESC) AS rn

FROM `twttr-recos-ml-prod.user\_events.valid\_user\_follows` F

LEFT JOIN tweet\_count T

ON F.destinationId=T.userId

WHERE DATE(F.\_PARTITIONTIME) = date\_latest\_follows

) SELECT \*, date\_latest\_tweet AS ds FROM all\_follows WHERE rn <= 2000

;