-- get latest partition of candidates with data

DECLARE date\_candidates DATE;

SET date\_candidates = (SELECT DATE(TIMESTAMP\_MILLIS($start\_time$)));

CREATE TABLE IF NOT EXISTS `twttr-recos-ml-prod.realgraph.candidates\_sampled` AS

SELECT \* FROM `twttr-recos-ml-prod.realgraph.candidates\_for\_training` LIMIT 100;

-- remove previous output snapshot (if exists) to avoid double-writing

DELETE

FROM `twttr-recos-ml-prod.realgraph.candidates\_sampled`

WHERE ds = date\_candidates;

-- sample from candidates table instead of recomputing features

INSERT INTO `twttr-recos-ml-prod.realgraph.candidates\_sampled`

SELECT \* FROM `twttr-recos-ml-prod.realgraph.candidates\_for\_training`

WHERE MOD(ABS(FARM\_FINGERPRINT(CONCAT(source\_id, '\_', destination\_id))), 100) = $mod\_remainder$

AND ds = date\_candidates;