from google.cloud import bigquery

client = bigquery.Client()

sql = """

WITH all\_journeys AS (

SELECT

dp024.channel\_visit\_id,

dp024.page,

dp024.page\_number,

dp024.digital\_id,

dp024.hit\_date\_time

FROM `bt-con-digital-dp-data.digital\_touchpoints\_rw.dp\_024\_ee\_pages\_viewed` AS dp024

WHERE DATE(dp024.hit\_date\_time) BETWEEN DATE '2024-12-13' AND DATE '2024-12-20'

AND LOWER(dp024.platform) = 'web'

),

springboard\_start AS (

SELECT DISTINCT channel\_visit\_id

FROM all\_journeys

WHERE LOWER(page) = 'springboard'

AND page\_number = 1

),

journey\_length AS (

SELECT

channel\_visit\_id,

COUNT(\*) AS length\_of\_journey

FROM all\_journeys

WHERE LOWER(page) NOT LIKE '%route%'

AND LOWER(page) NOT LIKE '%wrapper%'

GROUP BY 1

HAVING length\_of\_journey > 1

AND length\_of\_journey <= 15

)

SELECT dp013.\*, dp024.\*, aj.page AS journey\_page, aj.page\_number AS journey\_page\_number

FROM (

SELECT traceability.\*,

ROW\_NUMBER() OVER (PARTITION BY cdm\_join\_key ORDER BY request\_time DESC) AS last\_record

FROM `bt-bvp-ml-plat-ai-pipe-exp.hypothesis\_testing.dp013\_acquisition` AS traceability

WHERE LOWER(traceability.product\_type) = 'broadband'

AND DATE(request\_time) = DATE '2024-12-20'

AND cdm\_join\_key IS NOT NULL

)

QUALIFY last\_record = 1 AS dp013

INNER JOIN `bt-bvp-dataprod-dp-prod.cons\_bvp\_dp\_hhvm\_ro.vw\_hhv\_ibro\_subscription\_base` AS ibro

ON dp013.ee\_single\_id = ibro.single\_id

AND ibro.snapshot\_date = DATE(dp013.request\_time)

AND ibro.service = 'Mobile'

AND ibro.snapshot\_date = DATE '2024-12-20'

INNER JOIN `bt-con-coretelco-dp-data.cons\_dp\_account\_ro.avv\_ctd\_profile\_hub\_identity` AS id

ON ibro.single\_id = id.individual\_id

LEFT JOIN `bt-con-digital-dp-data.digital\_touchpoints\_rw.dp\_024\_ee\_pages\_viewed` AS dp024

ON id.digital\_id = dp024.digital\_id

AND TIMESTAMP(dp024.hit\_date\_time) <= dp013.request\_time

AND DATE(dp024.hit\_date\_time) BETWEEN DATE '2024-12-13' AND DATE '2024-12-20'

LEFT JOIN all\_journeys AS aj

ON dp024.channel\_visit\_id = aj.channel\_visit\_id

LEFT JOIN springboard\_start AS ss

ON aj.channel\_visit\_id = ss.channel\_visit\_id

LEFT JOIN journey\_length AS jl

ON aj.channel\_visit\_id = jl.channel\_visit\_id

WHERE LOWER(aj.page) NOT LIKE '%route%'

AND LOWER(aj.page) NOT LIKE '%wrapper%'

"""

# Run the query and convert to dataframe

df = client.query\_and\_wait(sql).to\_dataframe()