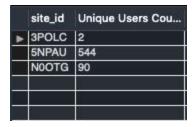
1. Consider only the rows with country_id = "BDV" (there are 844 such rows). For each site_id, we can compute the number of unique user_id's found in these 844 rows. Which site_id has the largest number of unique users? And what's the number?

Code:

SELECT site_id, COUNT(DISTINCT user_id) AS 'Unique Users Count' FROM sample WHERE country_id = 'BDV' GROUP BY site_id;

Output:



Answer: Site_id = '5NPAU' has the largest number of unique users with 544 unique users.

2. Between 2019-02-03 00:00:00 and 2019-02-04 23:59:59, there are four users who visited a certain site more than 10 times. Find these four users & which sites they (each) visited more than 10 times. (Simply provides four triples in the form (user_id, site_id, number of visits) in the box below.)

Code:

SELECT user_id, site_id, COUNT(user_id) AS 'Number of Visits'
FROM sample
WHERE ts BETWEEN '2019-02-03 00:00:00' AND '2019-02-04 23:59:59'
GROUP BY site_id, user_id
HAVING COUNT(user_id) > 10;

Output:

| user_id | site_id | Number of Vis |
|---------|----------------------------|---|
| LC3C7E | 3POLC | 15 |
| LC3A59 | N0OTG | 26 |
| LC06C3 | N0OTG | 25 |
| LC3C9D | N0OTG | 17 |
| | | |
| | | |
| | LC3C7E LC3A59 LC06C3 | user_id site_id LC3C7E 3POLC LC3A59 N0OTG LC06C3 N0OTG LC3C9D N0OTG |

Answer: (LC3C7E, 3POLC, 15), (LC3A59, N0OTG, 26), (LC06C3,N0OTG, 25), (LC3C9D, N0OTG, 17)

3. For each site, compute the unique number of users whose last visit (found in the original data set) was to that site. For instance, user "LC3561"'s last visit is to "N00TG" based on timestamp data. Based on this measure, what are top three sites? (hint: site "3POLC" is ranked at 5th with 28 users whose last visit in the data set was to 3POLC; simply provide three pairs in the form (site_id, number of users).)

Code:

SELECT sample1.site_id, COUNT(DISTINCT sample1.user_id) AS 'number of users' FROM sample AS sample1

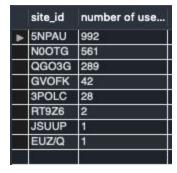
INNER JOIN (SELECT user_id, max(ts) AS last_ts

FROM sample

GROUP BY user_id) AS sample2

ON sample2.user_id=sample1.user_id AND sample2.last_ts = sample1.ts GROUP BY sample1.site_id ORDER BY COUNT(sample1.user_id) DESC;

Output:



Answer: (5NPAU, 992), (N00TG, 561), (QG03G, 289)

4. For each user, determine the first site he/she visited and the last site he/she visited based on the timestamp data. Compute the number of users whose first/last visits are to the same website. What is the number?

Code:

SELECT COUNT(last_sample.user_id) AS 'number of users' FROM

(SELECT sample1.user_id, sample1.ts, sample1.site_id FROM sample AS sample1 INNER JOIN (SELECT user_id, max(ts) AS last_ts FROM sample GROUP BY user_id) AS sample2

ON sample2.user_id = sample1.user_id AND sample2.last_ts = sample1.ts) AS last_sample
JOIN

(SELECT sample1.user_id, sample1.ts, sample1.site_id FROM sample AS sample1 INNER JOIN (SELECT user_id, min(ts) AS first_ts FROM sample GROUP BY user_id) AS sample3

ON sample3.user_id = sample1.user_id AND sample3.first_ts = sample1.ts) AS first_sample
ON last_sample.user_id = first_sample.user_id
WHERE last_sample.site_id = first_sample.site_id;

Output:



Answer: The number of users whose first/last visits are to the same website is 1670 (assuming that users with a single record also count)