

Lab 3 – Phân tích hiệu suất Website

Trong bài học này, các bạn sẽ đóng vai trò là một nhà phân tích dữ liệu để thực hiện phân tích lưu lượng website từ đó làm gia tăng hiệu quả tiếp thị.

1. Yêu cầu 1 – Tìm các trang web có lượt xem nhiều nhất

Một quản lý mới nhận chức muốn tìm các trang website có lượt xem nhiều nhất được xếp hạng theo số lượng session tính đến trước ngày 09-06-2012. Công việc của bạn là viết truy vấn để thực hiện yêu cầu này, sau đó đưa ra nhận xét về kết quả trả về.

Gợi ý:

- Sử dụng bảng website_pageviews
- Sử dụng Group by
- Kết quả trả về như hình dưới:

| pageview_url | pvs |
|---------------------------|-------|
| /home | 10322 |
| /products | 4194 |
| /the-original-mr-fuzzy | 3004 |
| /cart | 1293 |
| /shipping | 860 |
| /billing | 710 |
| /thank-you-for-your-order | 305 |

(pvs là viết tắt cho từ page view session, hãy sử dụng ‘as’)

2. Yêu cầu 2 – Tìm tổng số lượng session truy cập mỗi trang web vào lần đầu tiên

Ở yêu cầu này, người quản lý muốn xem có bao nhiêu session ghé thăm mỗi trang web vào lần đầu tiên khi bắt đầu session tính đến trước ngày 12-06-2012. Để dễ hiểu, các bạn sẽ xem ví dụ dưới đây:

Mỗi khi một người dùng truy cập website của bạn, sẽ tạo ra một phiên (session), trong một session người dùng có thể click qua rất nhiều trang web trên website của bạn, giả sử trang web của bạn đang có 3 trang web là a,b,c. Như vậy các bạn cần tính với mỗi trang a,b,c thì có bao nhiêu phiên mà thứ tự truy cập các trang này là thứ nhất. Bởi vì người dùng X có thể truy cập trang a trước sau đó với truy cập trang b, còn người dùng Y sẽ truy cập trang b trước sau đó với truy cập trang a.

Mục đích của yêu cầu này là người quản lý đang muốn xem trang web nào đang được người dùng truy cập đầu tiên nhiều nhất. Công việc của bạn là viết được câu truy vấn để thực hiện được yêu cầu của người quản lý và đưa ra nhận xét về kết quả trả về.

Gợi ý:

- Đối với yêu cầu này, các bạn sẽ phải viết hai câu truy vấn. Câu truy vấn đầu tiên để tìm web_id đầu tiên trong mỗi session. Câu thứ hai sẽ dựa vào web_id tìm được để tính tổng số lượng các pageview_url được truy cập vào lần đầu tiên.

- Gợi ý:

Câu truy vấn thứ nhất được viết như sau:

use mavenfuzzyfactory;

create temporary table first_pv_per_session

select

website_session_id,

min(website_pageview_id) as first_pv

from website_pageviews

where created_at < '2012-06-12'

group by website_session_id;

Ở câu truy vấn đầu tiên này, các bạn sẽ tạo ra một bảng tạm là first_pv_per_session để lưu kết quả.

Trong câu truy vấn thứ hai, các bạn sẽ cần sử dụng bảng này để join với bảng website_pageviews sau đó sẽ thực hiện group by theo pageview_url để tính tổng số lượng các pageview_url được truy cập vào lần đầu tiên.

- Kết quả sau khi chạy câu truy vấn thứ hai:

| landing_page_url | sessions_hitting_page |
|------------------|-----------------------|
| /home | 10614 |

3. Yêu cầu 3 – Phân tích landing page

Ở yêu cầu trước, bạn đã biết rằng tất cả lượng truy cập hiện đang tập chung vào trang chủ (/home) đầu tiên. Giờ người quản lý muốn bạn kiểm tra xem trang chủ đó đang hoạt động như thế nào. Người quản lý yêu cầu tìm tỷ lệ thoát phiên của trang chủ được thể hiện qua ba đại lượng session, bounced session (số session bị thoát phiên), và % bounced session

trên tổng số lượng session . Công việc của bạn là viết được câu truy vấn như người quản lý yêu cầu và đưa ra nhận xét đối với kết quả trả về.

Gợi ý:

- Hãy chạy các câu truy vấn dưới đây, sau đó dựa vào ghi chú của từng câu truy vấn để có thể viết được câu truy vấn thực hiện yêu cầu của người quản lý.

```
use mavenfuzzyfactory;
```

```
-- Bảng first_pageviews để tìm ra các website_id được truy cập đầu tiên
```

```
create temporary table first_pageviews
```

```
select
```

```
    website_session_id,
```

```
    min(website_pageview_id) as min_pageview_id
```

```
from website_pageviews
```

```
where created_at < '2012-06-14'
```

```
group by website_session_id;
```

```
create temporary table session_w_home_landing_page
```

```
select
```

```
    first_pageviews.website_session_id,
```

```
    website_pageviews.pageview_url as landing_page
```

```
from first_pageviews
```

```
    left join website_pageviews
```

```
        on website_pageviews.website_pageview_id =
```

```
first_pageviews.min_pageview_id
```

```
where website_pageviews.pageview_url = '/home';
```

```
-- bảng bounced_sessions để tìm ra các session chỉ chứa 1 một trang web là /home, nếu session chỉ truy cập duy nhất 1 trang là /home thì cũng đồng nghĩa session này đã xảy ra tình trạng thoát phiên
```

```

create temporary table bounced_sessions
select
    session_w_home_landing_page.website_session_id,
    session_w_home_landing_page.landing_page,
    count(website_pageviews.website_pageview_id) as count_of_pages_viewed
from session_w_home_landing_page
left join website_pageviews
    on website_pageviews.website_session_id =
session_w_home_landing_page.website_session_id
group by
    session_w_home_landing_page.website_session_id,
    session_w_home_landing_page.landing_page
having
    count(website_pageviews.website_pageview_id) = 1;

```

-- truy vấn này để tìm ra những phiên có trong bảng session_w_home_landing_page nhưng không có trong bảng bounced_sessions, những phiên có trong bảng bounced_sessions chính là những phiên xảy ra tình trạng thoát phiên

```

select
    session_w_home_landing_page.website_session_id,
    bounced_sessions.website_session_id as bounced_website_session_id
from session_w_home_landing_page
left join bounced_sessions
    on session_w_home_landing_page.website_session_id =
bounced_sessions.website_session_id
order by
    session_w_home_landing_page.website_session_id;

```

- Kết quả cuối cùng sẽ được thể hiện như hình dưới :

| sessions | bounced_sessions | bounced_rate |
|----------|------------------|--------------|
| 10942 | 6475 | 0.5918 |

4. Yêu cầu 4 – A/B testing

Dựa trên phân tích tỷ lệ thoát phiên của bạn, công ty đã chạy một trang đích tùy chỉnh mới (/lander-1), trang mới này và trang chủ cũ (/home) sẽ được chia đều 50/50 lưu lượng truy cập (nonbrand và gsearch). Đây còn được gọi là A/B testing. Giờ người quản lý muốn xem tỷ lệ thoát phiên của hai trang website mới và cũ. Và hãy đảm bảo chỉ xét trong khoảng thời gian mà /lander-1 đang có lưu lượng truy cập để giữ tính công bằng. (Chỉ lấy các phiên tạo trước ngày 28-07-2012, vì đây chính là ngày người quản lý gửi yêu cầu cho bạn)

Gợi ý:

Bước 1: Viết câu truy vấn để tìm ra ngày đầu tiên mà '/lander-1' được thêm vào:

select

min(created_at) as first_created_at,

min(website_pageview_id) as first_pageview_id

from website_pageviews

where pageview_url = '/lander-1' and created_at is not null;

Bước 2: lọc ra các web_pageview_id trong khoảng thời gian '/lander-1' đã được thêm

create temporary table first_test_pageviews

select

website_pageviews.website_session_id,

min(website_pageviews.website_pageview_id) as min_pageview_id

from website_pageviews

inner join website_sessions

```

on website_sessions.website_session_id =
website_pageviews.website_session_id

and website_sessions.created_at < '2012-07-28'

and website_pageviews.website_pageview_id > 23504

and utm_source = 'gsearch'

and utm_campaign = 'nonbrand'

group by

website_pageviews.website_session_id;

```

Bước 3: map web_pageview_id và landing page (/home, /lander-1)

create temporary table nonbrand_test_session_w_landing_page

select

```

first_test_pageviews.website_session_id,

website_pageviews.pageview_url as landing_page

from first_test_pageviews

left join website_pageviews

on website_pageviews.website_pageview_id =
first_test_pageviews.min_pageview_id

where website_pageviews.pageview_url in ('/home', '/lander-1');

```

Bước 4: Viết truy vấn lọc ra các web_session_id bị thoát phiên (Sinh viên thực hiện)

Bước 5: Viết truy vấn tính tỷ lệ thoát phiên của trang '/home', '/lander-1' (Sinh viên thực hiện), sau đó đưa ra nhận xét.

Kết quả cuối cùng có dạng như sau:

| | landing_page | sessions | bounced_sessions | bounce_rate |
|---|--------------|----------|------------------|-------------|
| ▶ | /home | 2227 | 1300 | 0.5837 |
| | /lander-1 | 2278 | 1210 | 0.5312 |

5. Yêu cầu 5

Sau khi đã biết landing page mới là /lander-1 đạt hiệu quả về mặt giảm tỷ lệ thoát phiên, giờ người quản lý muốn biết tổng tỷ lệ thoát phiên của toàn trang web theo từng tuần, bắt đầu từ ngày 01-06-2012 đến ngày 31-08-2012, trong kết quả trả về, người quản lý cũng muốn hiển thị luôn số phiên đang được điều hướng ở trang /home và số phiên đang được điều hướng sang trang /lander-1. Công việc của bạn là viết truy vấn giúp thực hiện yêu cầu của người quản lý.

Gợi ý:

Bước 1: Tạo bảng tạm để lọc các phiên thoả mãn điều kiện

create temporary table sessions_w_min_pv_id_and_view_count

select

website_sessions.website_session_id,

min(website_pageviews.website_pageview_id) as first_pageview_id,

count(website_pageviews.website_pageview_id) as count_pageviews

from website_sessions

left join website_pageviews

on website_sessions.website_session_id = website_pageviews.website_session_id

where website_sessions.created_at > '2012-06-01'

and website_sessions.created_at < '2012-08-31'

and website_sessions.utm_source = 'gsearch'

and website_sessions.utm_campaign = 'nonbrand'

group by

```
website_sessions.website_session_id;
```

Bước 2: Tạo bảng tạm để đếm số lượng count_pageviews trong mỗi phiên => mục đích để lọc ra các phiên bị thoát ở bước sau

```
create temporary table sessions_w_counts_lander_and_created_at
```

```
select
```

```
    sessions_w_min_pv_id_and_view_count.website_session_id,
```

```
    sessions_w_min_pv_id_and_view_count.first_pageview_id,
```

```
    sessions_w_min_pv_id_and_view_count.count_pageviews,
```

```
    website_pageviews.pageview_url as landing_page,
```

```
    website_pageviews.created_at as session_created_at
```

```
from sessions_w_min_pv_id_and_view_count
```

```
    left join website_pageviews
```

```
        on      sessions_w_min_pv_id_and_view_count.first_pageview_id      =  
        website_pageviews.website_pageview_id;
```

Bước 3: Hiển thị tỷ lệ thoát phiên trên toàn trang web theo tuần, đồng thời hiển thị cả số phiên được phân vào /home và số phiên được phân vào /lander-1 (Sinh viên thực hiện)

Kết quả cuối cùng có dạng như sau:

| | week_start_date | total_sessions | bounce_sessions | home_sessions | lander_sessions |
|---|-----------------|----------------|-----------------|---------------|-----------------|
| ► | 2012-06-01 | 229 | 0.60262 | 229 | 0 |
| | 2012-06-03 | 796 | 0.58920 | 796 | 0 |
| | 2012-06-10 | 869 | 0.61565 | 869 | 0 |
| | 2012-06-17 | 843 | 0.55991 | 512 | 331 |
| | 2012-06-24 | 754 | 0.58621 | 367 | 387 |
| | 2012-07-01 | 789 | 0.57795 | 401 | 388 |
| | 2012-07-08 | 794 | 0.56927 | 384 | 410 |
| | 2012-07-15 | 853 | 0.54045 | 428 | 425 |
| | 2012-07-22 | 788 | 0.50635 | 397 | 391 |
| | 2012-07-29 | 1028 | 0.50000 | 54 | 974 |
| | 2012-08-05 | 1080 | 0.54722 | 0 | 1080 |
| | 2012-08-12 | 992 | 0.50605 | 0 | 992 |
| | 2012-08-19 | 1022 | 0.50391 | 0 | 1022 |
| | 2012-08-26 | 765 | 0.54641 | 0 | 765 |