Analyzing Top Website Pages & Entry Pages

Task : Find the top website page before 2012-06-09.

```
SELECT
    pageview_url,
    COUNT(DISTINCT website_pageview_id) as pageview_count
FROM website_pageviews
WHERE created_at < '2012-06-09'
GROUP BY pageview_url
ORDER BY pageview_count DESC;
```

pageview_url	pageview_count
/home	10403
/products	4239
/the-original-mr-fuzzy	3037
/cart	1306
/shipping	869
/billing	716
/thank-you-for-your-order	306

The home webpage is the most visited page and the thank-you-for-your-order is the least visited page.

Task: Find the Top Entry/Landing Pages for the website.

Step 1: Each page that is viewed by a website visitor is given an identification. Determine the very first page id each website visitor sees before the date '2012-06-12'.

```
CREATE TEMPORARY TABLE first_pageviews
SELECT
    website_session_id,
    MIN(website_pageview_id) AS min_pageview_id
FROM website_pageviews
WHERE created_at < '2012-06-12'
GROUP BY website_session_id;
```

website_session_id	min_pageview_id
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	12
9	14
10	15

Step 2: Find the URL of the page associated with these page view identifications that the customer saw..

```
SELECT
    w.pageview_url AS landing_page,
    COUNT(f.website_session_id) AS sessions_hit_landing_page
FROM first_pageviews f
LEFT JOIN website_pageviews w ON w.website_pageview_id = f.min_pageview_id
GROUP BY w.pageview_url;
```

landing_page	sessions_hit_landing_page
/home	11048

The home page is the top entry page of the website at this time as all the visitors land on this page. The performance of the homepage should be analyzed and optimized so that the experience of the homepage can be as good as possible for all first time visitors of the website.

Analyzing Bounce Rates and Landing Page Tests

Task: Determine the bounce rates for traffic landing on the homepage? Provide the session count, bounced sessions, and percent of sessions which bounced.

Step 1: Find all the first website pageview identifications for visitors who landed on the home page.

```
CREATE TEMPORARY TABLE first_pageviews_demo
SELECT
    wp.website_session_id,
    MIN(wp.website_pageview_id)AS min_pageview_id
FROM website_pageviews wp
INNER JOIN website_sessions ws
ON ws.website_session_id = wp.website_session_id
AND ws.created_at BETWEEN'2014-01-01' AND '2014-02-01'
GROUP BY wp.website_session_id;
```

website_session_id	min_pageview_id
175252	415126
175253	415128
175254	415134
175255	415136
175256	415141
175257	415145
175258	415149
175259	415153
175260	415154
175261	415155

STEP 2: Associating each first pageview id with the appropriate webpage.

```
CREATE TEMPORARY TABLE session_w_landing_page_demo
SELECT
    fpd.website_session_id,
    wp.pageview_url AS landing_page
FROM first_pageviews_demo fpd
LEFT JOIN website_pageviews wp
ON wp.website_pageview_id = fpd.min_pageview_id;
```

website_session_id	landing_page
175252	/lander-2
175253	/home
175254	/lander-2
175255	/home
175256	/lander-2
175257	/lander-3
175258	/lander-2
175259	/home
175260	/lander-2
175261	/lander-3

STEP 3: Counting the amount of pages each visitor has seen.

```
SELECT
    swlpd.website_session_id,
    swlpd.landing_page,
    COUNT(wp.website_pageview_id)AS pageviews_cnt

FROM session_w_landing_page_demo swlpd
LEFT JOIN website_pageviews wp
ON wp.website_session_id = swlpd.website_session_id
```

website_session_id	landing_page	pageviews_cnt
175252	/lander-2	3
175253	/home	3
175254	/lander-2	6
175255	/home	7
175256	/lander-2	1
175257	/lander-3	3
175258	/lander-2	2
175259	/home	1
175260	/lander-2	1
175261	/lander-3	1
175262	/lander-2	3
175263	/home	4
175264	/lander-2	1
175265	/home	1
175266	/lander-2	1
175267	/lander-2	3
175268	/home	7

STEP 4: Further filtering this table to show visitors who bounced

```
DROP TABLE IF EXISTS bounced_session_only;

CREATE TEMPORARY TABLE bounced_session_only
SELECT
    swlpd.website_session_id,
    swlpd.landing_page,
    COUNT(wp.website_pageview_id)AS pageviews_cnt

FROM session_w_landing_page_demo swlpd
LEFT JOIN website_pageviews wp
ON wp.website_session_id = swlpd.website_session_id
GROUP BY 1,2
HAVING COUNT(wp.website_pageview_id) = 1;

SELECT * FROM bounced_session_only
LIMIT 5;
```

website_session_id	landing_page	pageviews_cnt
175256	/lander-2	1
175259	/home	1
175260	/lander-2	1
175261	/lander-3	1
175264	/lander-2	1

Step 5: Summarizing by counting total sessions and bounced sessions, by landing page.

```
SELECT
    swlpd.landing_page,
    COUNT(DISTINCT swlpd.website_session_id)AS sessions,
    COUNT(DISTINCT bso.website_session_id) AS bounced_sessions,
    COUNT(DISTINCT bso.website_session_id)/COUNT(DISTINCT swlpd.website_session_id)AS bounced_rate
FROM session_w_landing_page_demo swlpd
LEFT JOIN bounced_session_only bso
ON swlpd.website_session_id = bso.website_session_id
GROUP BY 1;
```

landing_page	sessions	bounced_sessions	bounced_rate
/home	4093	1575	0.3848
/lander-2	6500	2855	0.4392
/lander-3	4232	2606	0.6158

Calculate Bounce Rates

Task: Determine the bounce rate for traffic landing on the homepage

STEP 1: finding the first website_pageview_id for relevant sessions

```
DROP TABLE IF EXISTS first_pageviews;

CREATE TEMPORARY TABLE first_pageviews
SELECT
    website_session_id,
    MIN(website_pageview_id) AS min_pageview_id
FROM website_pageviews
WHERE created_at < '2012-06-12'
GROUP BY website_session_id;

SELECT * FROM first_pageviews
LIMIT 5;
```

website_session_id	min_pageview_id
1	1
2	2
3	3
4	4
5	5

STEP 2: identifying the landing page of each session

```sql
DROP TABLE IF EXISTS session_w_landing_page_demo;

CREATE TEMPORARY TABLE session_w_landing_page_demo
SELECT
    fpd.website_session_id,
    wp.pageview_url AS landing_page
FROM first_pageviews_demo fpd
LEFT JOIN website_pageviews wp
ON wp.website_pageview_id = fpd.min_pageview_id # website pageview is the landing view
WHERE wp.pageview_url = '/home';

SELECT * FROM session_w_landing_page_demo
LIMIT 5;
```

website_session_id	landing_page
175253	/home
175255	/home
175259	/home
175263	/home
175265	/home

STEP 3: counting pageviews for each session, to identify "bounces"

```sql
DROP TABLE IF EXISTS bounced_session_only;

CREATE TEMPORARY TABLE bounced_session_only
SELECT
    swlpd.website_session_id,
    swlpd.landing_page,
    COUNT(wp.website_pageview_id)AS pageviews_cnt

FROM session_w_landing_page_demo swlpd
LEFT JOIN website_pageviews wp
ON wp.website_session_id = swlpd.website_session_id
GROUP BY 1,2
HAVING COUNT(wp.website_pageview_id) = 1;

SELECT * FROM bounced_session_only
LIMIT 5;
```

website_session_id	landing_page	pageviews_cnt
175259	/home	1
175265	/home	1
175270	/home	1
175280	/home	1
175285	/home	1

Step 4: summarizing by counting total sessions and bounced sessions, by landing page

```
SELECT
    swlpd.landing_page,
    COUNT(DISTINCT swlpd.website_session_id)AS sessions,
    COUNT(DISTINCT bso.website_session_id) AS bounced_sessions,
    COUNT(DISTINCT bso.website_session_id)/COUNT(DISTINCT swlpd.website_session_id)AS bounced_rate
FROM session_w_landing_page_demo swlpd
LEFT JOIN bounced_session_only bso
ON swlpd.website_session_id = bso.website_session_id
GROUP BY 1;
```

landing_page	sessions	bounced_sessions	bounced_rate
/home	4093	1575	0.3848

Analysis : This is a significantly high bounce rate. The web development team will setup another landing page and perform an A/B split test with this new page against the home page to try and improve the bounce rate performance.

Analyzing Landing Page Test

```
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;
```

first_created_at	first_pageview_id
2012-06-19 00:35:54	23504

Step 2: find the first website_pageview_id for relevant session

```
DROP TABLE IF EXISTS first_test_pageviews;

CREATE TEMPORARY TABLE first_test_pageviews
SELECT
    wp.website_session_id,
    MIN(wp.website_pageview_id) AS min_pageview_id
FROM website_pageviews wp
INNER JOIN website_sessions ws
        ON ws.website_session_id = wp.website_session_id
       AND ws.created_at < '2012-07-28' #prescribed by the assignment
       AND wp.website_pageview_id > 23504 # the min_pageview_id we found for
       AND utm_source = 'gsearch'
       AND utm_campaign = 'nonbrand'
GROUP BY wp.website_session_id;

SELECT * FROM first_test_pageviews LIMIT 5;
```

website_session_id	min_pageview_id
11684	23505
11685	23506
11686	23507
11687	23509
11688	23510

\

Step 3: identifying the landing page of each session

```
DROP TABLE IF EXISTS nonbrand_test_sessions_w_landing_page;

CREATE TEMPORARY TABLE nonbrand_test_sessions_w_landing_page
SELECT
    ftp.website_session_id,
    wp.pageview_url AS landing_page
FROM first_test_pageviews ftp
LEFT JOIN website_pageviews wp
ON wp.website_pageview_id = ftp.min_pageview_id
WHERE wp.pageview_url IN('/home','/lander-1');

SELECT * FROM nonbrand_test_sessions_w_landing_page LIMIT 5;
```

website_session_id	landing_page
11684	/home
11685	/lander-1
11686	/lander-1
11687	/home
11688	/home

Step 4: counting pageviews for each session, to identify "bounce"

```
DROP TABLE IF EXISTS nonbrand_test_bounced_sessions;

CREATE TEMPORARY TABLE nonbrand_test_bounced_sessions
SELECT
    ntswlp.website_session_id,
    ntswlp.landing_page,
    COUNT(wp.website_pageview_id) AS pageviewed_cnt
FROM nonbrand_test_sessions_w_landing_page ntswlp
LEFT JOIN website_pageviews wp
ON ntswlp.website_session_id = wp.website_session_id
GROUP BY 1,2
HAVING COUNT(wp.website_pageview_id) = 1;

SELECT * FROM nonbrand_test_bounced_sessions LIMIT 5;
```

website_session_id	landing_page	pageviewed_cnt
11684	/home	1
11685	/lander-1	1
11687	/home	1
11688	/home	1
11690	/home	1

Step 5: summarizing total sessions and bounced sessions, by LP

```
SELECT
    ntswlp.landing_page,
    ntswlp.website_session_id,
    ntbs.website_session_id AS bounced_website_session_id
FROM nonbrand_test_sessions_w_landing_page ntswlp
LEFT JOIN nonbrand_test_bounced_sessions ntbs
ON ntswlp.website_session_id = ntbs.website_session_id
ORDER BY ntswlp.website_session_id;
```

landing_page	website_session_id	bounced_website_session_id
/home	11684	11684
/lander-1	11685	11685
/lander-1	11686	NULL
/home	11687	11687
/home	11688	11688
/lander-1	11689	NULL
/home	11690	11690
/lander-1	11691	NULL
/lander-1	11692	11692
/lander-1	11693	NULL

```
SELECT
    ntswlp.landing_page,
    COUNT(DISTINCT ntswlp.website_session_id) AS sessions,
    COUNT(DISTINCT ntbs.website_session_id) AS bounced_sessions,
    COUNT(DISTINCT ntbs.website_session_id)/COUNT(DISTINCT ntswlp.website_session_id) AS bounced_rate
FROM nonbrand_test_sessions_w_landing_page ntswlp
LEFT JOIN nonbrand_test_bounced_sessions ntbs
ON ntswlp.website_session_id = ntbs.website_session_id
GROUP BY 1;
```

landing_page	sessions	bounced_sessions	bounced_rate
/home	2261	1319	0.5834
/lander-1	2315	1232	0.5322

Analysis: The new landing page has a significantly lower bounce rate. This new page will continue to be used and the trends for this page will be monitored.

Landing Page Trend Analysis

Task: Determine the volume of paid search nonbrand traffic landing on /home and /lander-1, trended weekly since June 1st.Determine the overall paid search bounce rate trended weekly?

Step 1: finding the first website_pageview_id for relevant sessions

```
DROP TABLE IF EXISTS session_w_min_pv_id_and_view_cnt;

CREATE TEMPORARY TABLE session_w_min_pv_id_and_view_cnt
SELECT
    ws.website_session_id,
    MIN(wp.website_pageview_id) AS first_pageview_id,
    COUNT(wp.website_pageview_id) AS pageview_cnt

FROM website_sessions ws
LEFT JOIN website_pageviews wp
        ON ws.website_session_id = wp.website_session_id

WHERE ws.created_at > '2012-06-01'  # asked by requestor
    AND ws.created_at < '2012-08-31' # prescribed by the assignment
    AND utm_source = 'gsearch'
    AND utm_campaign = 'nonbrand'

GROUP BY 1;

SELECT * FROM session_w_min_pv_id_and_view_cnt LIMIT 5;
```

website_session_id	first_pageview_id	pageview_cnt
9350	18598	3
9351	18600	3
9352	18601	4
9354	18611	1
9356	18616	6

Step 2: identifying the landing page of each session

```
DROP TABLE IF EXISTS session_w_cnt_lander_and_created_at;

CREATE TEMPORARY TABLE session_w_cnt_lander_and_created_at
SELECT
    swmpiavc.website_session_id,
    swmpiavc.first_pageview_id,
    swmpiavc.pageview_cnt,
    wp.pageview_url AS landing_page,
    wp.created_at AS session_created_at

FROM session_w_min_pv_id_and_view_cnt swmpiavc
LEFT JOIN website_pageviews wp ON swmpiavc.first_pageview_id = wp.website_pageview_id;

SELECT * FROM session_w_cnt_lander_and_created_at LIMIT 5;
```

website_session_id	first_pageview_id	pageview_cnt	landing_page	session_created_at
9350	18598	3	/home	2012-06-01 00:05:11
9351	18600	3	/home	2012-06-01 00:06:39
9352	18601	4	/home	2012-06-01 00:08:27
9354	18611	1	/home	2012-06-01 01:08:43
9356	18616	6	/home	2012-06-01 01:37:31

Step 3: counting pageviews for each session, to identify "bounce"

```
SELECT
    YEARWEEK(session_created_at) AS year_week,
    MIN(DATE(session_created_at)) AS week_start_date,
    COUNT(DISTINCT website_session_id) AS total_sessions,
    COUNT(DISTINCT CASE WHEN pageview_cnt = 1 THEN website_session_id ELSE NULL END)AS bounce_session,
    COUNT(DISTINCT CASE WHEN pageview_cnt = 1 THEN website_session_id ELSE NULL END)*1.0/COUNT(DISTINCT website_session_id)AS bounce_rate,
    COUNT(DISTINCT CASE WHEN landing_page = '/home' THEN website_session_id ELSE NULL END) AS home_sessions,
    COUNT(DISTINCT CASE WHEN landing_page = '/lander-1' THEN website_session_id ELSE NULL END) AS lander_sessions
FROM session_w_cnt_lander_and_created_at
GROUP BY 1;
```

website_session_id	first_pageview_id	pageview_cnt	landing_page	session_created_at
9350	18598	3	/home	2012-06-01 00:05:11
9351	18600	3	/home	2012-06-01 00:06:39
9352	18601	4	/home	2012-06-01 00:08:27
9354	18611	1	/home	2012-06-01 01:08:43
9356	18616	6	/home	2012-06-01 01:37:31

Step 4: summarizing by week (bounce rate, sessions to each lander)

```
SELECT
    MIN(DATE(session_created_at)) AS week_start_date,
    COUNT(DISTINCT CASE WHEN pageview_cnt = 1 THEN website_session_id ELSE NULL END)*1.0/COUNT(DISTINCT website_session_id)AS bounce_rate,
    COUNT(DISTINCT CASE WHEN landing_page = '/home' THEN website_session_id ELSE NULL END) AS home_sessions,
    COUNT(DISTINCT CASE WHEN landing_page = '/lander-1' THEN website_session_id ELSE NULL END) AS lander_sessions

FROM session_w_cnt_lander_and_created_at

GROUP BY YEARWEEK(session_created_at) ;
```

week_start_date	bounce_rate	home_sessions	lander_sessions
2012-06-01	0.60571	175	0
2012-06-03	0.58712	792	0
2012-06-10	0.61600	875	0
2012-06-17	0.55819	492	350
2012-06-24	0.58278	369	386
2012-07-01	0.58205	392	388
2012-07-08	0.56679	390	411
2012-07-15	0.54235	429	421
2012-07-22	0.51382	402	394
2012-07-29	0.49708	33	995
2012-08-05	0.53818	0	1087
2012-08-12	0.51403	0	998
2012-08-19	0.50099	0	1012
2012-08-26	0.53782	0	833

Analysis: It seems that before 2012-06-10 only the home page was being used, then between 2012-06-17 to 2012-07-29 both pages were being used, and after 2012-08-05 only the new landing page was used. This change to the new landing page has caused an overall decrease in bounce rate which is a good improvement.