WEB TRAFFIC ANALYSIS USING MYSQL

INTRODUCTION :

You've just been hired as an eCommerce Database Analyst for Maven Fuzzy Factory, an online SITUATION retailer which has just launched their first product. As a member of the startup team, you will work with the CEO, the Head of Marketing, and the Website Manager to help steer the business. You will analyze and optimize marketing channels, measure and test website conversion performance, and use data to understand the impact of new product launches.

OBJECTIVE :

Analyze and optimize the business' marketing channels, website, and product portfolio for an online retailer.

DATA INFORMATION:

SQL file was provided my Maven analytics and executed on MySQL workbench to create different tables namely :

i) website_sessions

ii) website_pageviews

iii) orders

iv) products

v) order_items

vi) order-item_refunds

I will be working with the above six related tables, which contain eCommerce data about:

- Website Activity
- Products
- Orders and Refunds

I will use MySQL to understand how customers access and interact with the site, analyze landing page performance and conversion, and explore product-level sales

WEB TRAFFIC ANALYSIS USING MYSQL

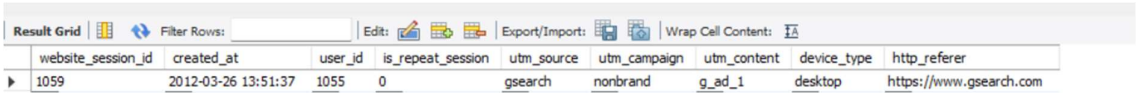
ANALYSIS

1) We will perform traffic analysis & optimization to Identify top traffic sources, measure their conversion rates, analyze trends, and use segmentation for bidding optimization.

For this I will focus on the WEBSITE SESSIONS, PAGEVIEWS & ORDERS tables to see what are the data columns we have .

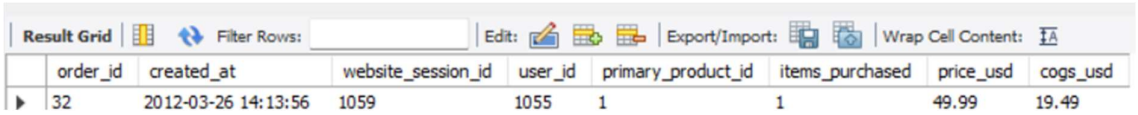
```
SELECT * FROM website_sessions WHERE website_session_id = 1059;
SELECT * FROM orders WHERE website_session_id = 1059;
SELECT * FROM website_pageviews WHERE website_session_id = 1059;
```

Website_sessions:



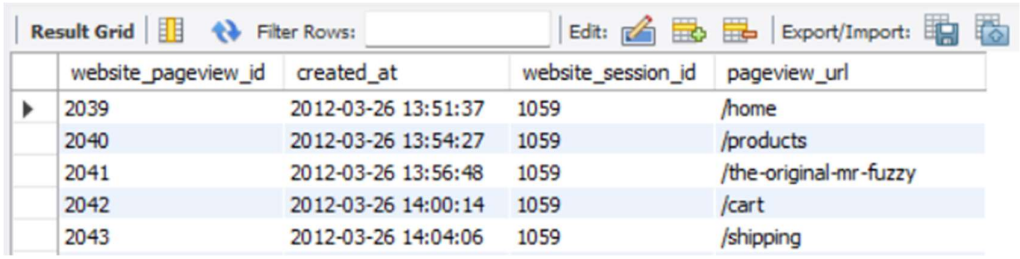
website_session_id	created_at	user_id	is_repeat_session	utm_source	utm_campaign	utm_content	device_type	http_referer
1059	2012-03-26 13:51:37	1055	0	gsearch	nonbrand	g_ad_1	desktop	https://www.gsearch.com

Orders:



order_id	created_at	website_session_id	user_id	primary_product_id	items_purchased	price_usd	cogs_usd
32	2012-03-26 14:13:56	1059	1055	1	1	49.99	19.49

Website_pageviews:

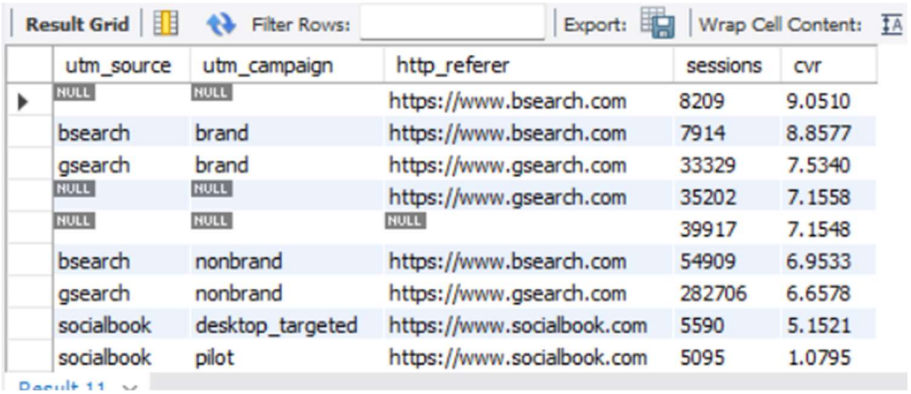


website_pageview_id	created_at	website_session_id	pageview_url
2039	2012-03-26 13:51:37	1059	/home
2040	2012-03-26 13:54:27	1059	/products
2041	2012-03-26 13:56:48	1059	/the-original-mr-fuzzy
2042	2012-03-26 14:00:14	1059	/cart
2043	2012-03-26 14:04:06	1059	/shipping

WEB TRAFFIC ANALYSIS USING MYSQL

Lets check where the bulk of our website sessions are coming from:

```
SELECT
    t1.utm_source,
    t1.utm_campaign,
    t1.http_referer,
    COUNT(DISTINCT t1.website_session_id) AS sessions,
    (COUNT(DISTINCT t2.order_id) / COUNT(DISTINCT t1.website_session_id)) * 100 AS cvr
FROM
    website_sessions t1
        LEFT JOIN
    orders t2 ON t1.website_session_id = t2.website_session_id
WHERE
    t1.created_at < '2012-04-12'
GROUP BY t1.utm_source , t1.utm_campaign , t1.http_referer
ORDER BY cvr DESC;
```



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

utm_source	utm_campaign	http_referer	sessions	cvr
NULL	NULL	https://www.bsearch.com	8209	9.0510
bsearch	brand	https://www.bsearch.com	7914	8.8577
gsearch	brand	https://www.gsearch.com	33329	7.5340
NULL	NULL	https://www.gsearch.com	35202	7.1558
NULL	NULL	NULL	39917	7.1548
bsearch	nonbrand	https://www.bsearch.com	54909	6.9533
gsearch	nonbrand	https://www.gsearch.com	282706	6.6578
socialbook	desktop_targeted	https://www.socialbook.com	5590	5.1521
socialbook	pilot	https://www.socialbook.com	5095	1.0795

Most sessions to orders conversion rate is coming from direct bsearch url with 9.05 conversion rate .

But the number of sessions is highest in gsearch non brand utm parameters.

Lets check if the huge number of gsearch non brand sessions are even driving sales ?

WEB TRAFFIC ANALYSIS USING MYSQL

```sql
SELECT
    COUNT(DISTINCT t1.website_session_id) AS sessions,
    COUNT(DISTINCT t2.order_id) AS orders,
    (COUNT(DISTINCT t2.order_id) / COUNT(DISTINCT t1.website_session_id)) * 100 AS cvr
FROM
    website_sessions t1
        LEFT JOIN
    orders t2 ON t1.website_session_id = t2.website_session_id
WHERE
        utm_source = 'gsearch'
        AND utm_campaign = 'nonbrand'
ORDER BY cvr DESC;
```

Result Grid | Filter Rows:

sessions	orders	cvr
282706	18822	6.6578

cvr = conversion rate

It seems to be a good conversion rate but can be improved taking into account the huge traffic we are getting .It has much more potential.

Lets check sessions to order conversion rate by device type :

```sql
SELECT
    t1.device_type,
    COUNT(DISTINCT t1.website_session_id) AS sessions,
    COUNT(DISTINCT t2.order_id) AS orders,
    (COUNT(DISTINCT t2.order_id) / COUNT(DISTINCT t1.website_session_id)) * 100 AS cvr
FROM
    website_sessions t1
        LEFT JOIN
    orders t2 ON t1.website_session_id = t2.website_session_id
GROUP BY device_type;
```

Result Grid | Filter Rows:

device_type	sessions	orders	cvr
desktop	327027	27805	8.5024
mobile	145844	4508	3.0910

Desktop has significantly higher sessions , orders and conversion rate as compared to the mobile devices with 8.5% conversion rate and mobile devices has just 3% rate.

WEB TRAFFIC ANALYSIS USING MYSQL

2) Now we will perform website content analysis by analyzing the most-visited pages and top entry pages, calculate bounce rates, build conversion funnels. Website content analysis is about understanding which pages are seen the most by your users, to identify where to focus on improving your business.

Lets check the most viewed web pages ranked by sessions:

```
SELECT
    pageview_url,
    COUNT(DISTINCT (website_session_id)) AS sessions
    FROM
    website_pageviews
GROUP BY pageview_url
ORDER BY 2 DESC;
```

Result Grid | Filter Rows: | Export:

pageview_url	sessions
/products	261231
/the-original-mr-fuzzy	162525
/home	137576
/lander-2	131170
/cart	94953
/lander-3	79000
/lander-5	68166
/shipping	64484
/billing-2	48441
/lander-1	47574
/thank-you-for-your-...	32313
/the-forever-love-bear	26033
/the-birthday-sugar-...	19046

The products page has the most sessions followed with /the-original-mr-fuzzy and /home .

3)Now we will perform Landing page trend analysis in which we will check the performance of your key landing pages .

Steps :

-- step1 finding the first website_pageview_id for relevant sessions

-- step2 identifying the landing page for each session

-- step 3 counting pageviews for each session, to identify "bounces”

-- step 4 summarizing by bounce rate and sessions

WEB TRAFFIC ANALYSIS USING MYSQL

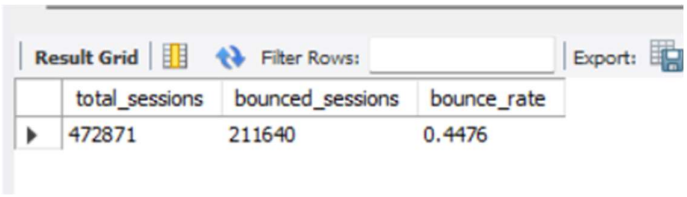
```sql
-- step1 finding the first website_pageview_id for relevant sessions
create temporary table sessions_w_min_pv_id_and_view_count1
SELECT
    t1.website_session_id,
    MIN(t2.website_pageview_id) as first_pageview_id,
    COUNT(t2.website_pageview_id) as count_pageviews
FROM
    website_sessions t1
        LEFT JOIN
     website_pageviews t2 ON t1.website_session_id = t2.website_session_id
GROUP BY t1.website_session_id;

-- step2 identifying the landing page for each session
select * from  sessions_w_min_pv_id_and_view_count1;
CREATE TEMPORARY TABLE sessions_w_counts_lander_and_created_at1
SELECT
    t1.website_session_id,
    t1.first_pageview_id,
    t1.count_pageviews,
    t2.pageview_url AS landing_page,
    t2.created_at as session_created_at
FROM
    sessions_w_min_pv_id_and_view_count1 t1
        LEFT JOIN

    website_pageviews t2 ON t1.first_pageview_id = t2.website_pageview_id;

-- step 3 &4 counting pageviews for each session, to identify "bounces"
select
count(distinct website_session_id) as total_sessions,
count(distinct case when count_pageviews = 1 then website_session_id else null end)
as bounced_sessions,
(count(distinct case when count_pageviews = 1 then website_session_id else null end)
)/(count(distinct website_session_id)) as bounce_rate
from sessions_w_counts_lander_and_created_at1
;
```

WEB TRAFFIC ANALYSIS USING MYSQL



Result Grid | Filter Rows: | Export:

total_sessions	bounced_sessions	bounce_rate
472871	211640	0.4476

There is 0.4476 bounce rate overall.

4) Conversion Funnels Analysis: I will perform conversion funnel analysis, I will look at each step in our conversion flow to see how many customers drop off and how many continue on at each step.

```
-- CONVERSION FUNNELS
-- We will create temporary tables using pageview data in order to build our multi-step funnels
-- We will first identify the sessions we care about, then bring in the relevant pageviews,
-- then flag each session as having made it to certain funnel steps, and finally perform a summary analysis.

-- Building Conversion Funnels

-- Business context
  -- we want to build a mini conversion funnel, from /lander-2 to /cart
  -- we want to know how many people reach each step and also dropoff rates
  -- for simplicity of the demo, we're looking at /lander-2 traffic only
  -- for simplicity of the demo , we're looking at customers who like Mr fuzzy only


-- STEP 1 : select all pageviews for relevant sessions
-- STEP 2 : identify each relevant pageview as the specific funnel step
-- STEP 3 : create the session-level conversion funnel view
-- STEP 4 : aggregate the data to assess funnel performance

  SELECT ws.website_session_id, wp.pageview_url, wp.created_at ,
  CASE WHEN pageview_url = '/products' THEN 1 ELSE 0 END AS products_page,
  CASE WHEN pageview_url = '/the-original-mr-fuzzy' THEN 1 ELSE 0 END AS mrfuzzy_page,
  CASE WHEN pageview_url = '/cart' THEN 1 ELSE 0 END AS  cart_page
  FROM website_sessions ws LEFT JOIN website_pageviews wp on ws.website_session_id= wp.website_session_id
  where ws.created_at BETWEEN '2014-01-01' AND '2014-02-01' AND
  wp.pageview_url IN ('/lander-2','/products','/the-original-mr-fuzzy','/cart')
  order by
  ws.website_session_id , wp.created_at;


  -- next we will put the previous query inside a subquery (simlar to temporary tables)
  -- we will group by website_session_id and take MAX() of each of the flags
  -- this MAX() becomes a made_it flag for that session, to show the session made it there

  select website_session_id, MAX(products_page) as products_made_it,
  MAX(mrfuzzy_page) as mrfuzzy_made_it  ,MAX(cart_page) as cart_made_it
  from (SELECT ws.website_session_id, wp.pageview_url, wp.created_at ,
  CASE WHEN pageview_url = '/products' THEN 1 ELSE 0 END AS products_page,
  CASE WHEN pageview_url = '/the-original-mr-fuzzy' THEN 1 ELSE 0 END AS mrfuzzy_page,
  CASE WHEN pageview_url = '/cart' THEN 1 ELSE 0 END AS  cart_page
```

WEB TRAFFIC ANALYSIS USING MYSQL

```
FROM website_sessions ws LEFT JOIN website_pageviews wp on ws.website_session_id= wp.website_session_id
where ws.created_at BETWEEN '2014-01-01' AND '2014-02-01' AND
wp.pageview_url IN ('/lander-2','/products','/the-original-mr-fuzzy','/cart')
order by
ws.website_session_id , wp.created_at) as pageview_level
group by website_session_id;

-- next,we will turn it into a temp table

#CREATE view session_level_made_it_flags_demo as
select website_session_id, MAX(products_page) as products_made_it,
MAX(mrfuzzy_page) as mrfuzzy_made_it  ,MAX(cart_page) as cart_made_it
from (SELECT ws.website_session_id, wp.pageview_url, wp.created_at ,
CASE WHEN pageview_url = '/products' THEN 1 ELSE 0 END AS products_page,
CASE WHEN pageview_url = '/the-original-mr-fuzzy' THEN 1 ELSE 0 END AS mrfuzzy_page,
CASE WHEN pageview_url = '/cart' THEN 1 ELSE 0 END AS  cart_page
FROM website_sessions ws LEFT JOIN website_pageviews wp on ws.website_session_id= wp.website_session_id
where ws.created_at BETWEEN '2014-01-01' AND '2014-02-01' AND
wp.pageview_url IN ('/lander-2','/products','/the-original-mr-fuzzy','/cart')
order by
ws.website_session_id , wp.created_at) as pageview_level
group by website_session_id;
select * from session_level_made_it_flags_demo;
```

```
-- taking count of total sessions and sessions making it to each pages
select COUNT(distinct website_session_id ) as sessions,
count(distinct case when products_made_it = 1 THEN website_session_id else NULL END) as products_count ,
count(distinct case when mrfuzzy_made_it = 1 THEN website_session_id else NULL END) as mrfuzzy_count,
count(distinct case when cart_made_it = 1 THEN website_session_id else NULL END) as cart_count
from session_level_made_it_flags_demo;

-- translating those counts to click rates for final ouput
select COUNT(distinct website_session_id ) as sessions,
count(distinct case when products_made_it = 1 THEN website_session_id else NULL END)/COUNT(DISTINCT website_session_id ) as lander_clickthrough_rate,
count(distinct case when mrfuzzy_made_it = 1 THEN website_session_id else NULL END)/COUNT(DISTINCT CASE WHEN products_made_it = 1 THEN website_session_id else NULL END) as products_clickthrough_rate,
count(distinct case when cart_made_it = 1 THEN website_session_id else NULL END)/count(distinct case when mrfuzzy_made_it = 1 THEN website_session_id else NULL END) as mr_fuzzy_clickthrough_rate
from session_level_made_it_flags_demo;
select distinct pageview_url from website_pageviews where website_session_id= 1059;

-- Assignment all steps funnel Aug 5th to Sept 5th

SELECT ws.website_session_id, wp.pageview_url, wp.created_at ,
CASE WHEN pageview_url = '/products' THEN 1 ELSE 0 END AS products_page,
CASE WHEN pageview_url = '/the-original-mr-fuzzy' THEN 1 ELSE 0 END AS mrfuzzy_page,
CASE WHEN pageview_url = '/cart' THEN 1 ELSE 0 END AS  cart_page,
CASE WHEN pageview_url = '/shipping' THEN 1 ELSE 0 END AS  shipping_page,
CASE WHEN pageview_url = '/billing' THEN 1 ELSE 0 END AS  billing_page,
CASE WHEN pageview_url = '/thank-you-for-your-order' THEN 1 ELSE 0 END AS  thankyou_page
```

WEB TRAFFIC ANALYSIS USING MYSQL

```
FROM website_sessions ws LEFT JOIN website_pageviews wp on ws.website_session_id= wp.website_session_id
where ws.utm_source = 'gsearch' and ws.utm_campaign= 'nonbrand'

order by
ws.website_session_id , wp.created_at) as pageview_level
group by website_session_id;

select COUNT(distinct website_session_id ) as sessions,
count(distinct case when products_made_it = 1 THEN website_session_id else NULL END) as products_count ,
count(distinct case when mrfuzzy_made_it = 1 THEN website_session_id else NULL END) as mrfuzzy_count,
count(distinct case when cart_made_it = 1 THEN website_session_id else NULL END) as cart_count,
count(distinct CASE WHEN shipping_made_it = 1 THEN website_session_id ELSE NULL END) as shipping_count,
count(distinct CASE WHEN billing_made_it = 1 THEN website_session_id ELSE NULL END) AS  billing_count,
count(distinct CASE WHEN thankyou_made_it = 1 THEN website_session_id ELSE NULL END) AS  thankyou_count
from session_level_made_it_flags_assignment2;

-- click rates
#create view session_made_it as
select COUNT(distinct website_session_id ) as sessions,
count(distinct case when products_made_it = 1 THEN website_session_id else NULL END)/COUNT(distinct website_session_id ) as lander_click_rt ,
count(distinct case when mrfuzzy_made_it = 1 THEN website_session_id else NULL END)/count(distinct case when products_made_it = 1 THEN website_session_id else NULL END) as products_click_rt,
count(distinct case when cart_made_it = 1 THEN website_session_id else NULL END)/count(distinct case when mrfuzzy_made_it = 1 THEN website_session_id else NULL END) as mrfuzzy_click_rt,
count(distinct CASE WHEN shipping_made_it = 1 THEN website_session_id ELSE NULL END)/count(distinct case when cart_made_it = 1 THEN website_session_id else NULL END) as cart_click_rt,
count(distinct CASE WHEN billing_made_it = 1 THEN website_session_id ELSE NULL END)/count(distinct CASE WHEN shipping_made_it = 1 THEN website_session_id ELSE NULL END) AS  shipping_click_rt,
count(distinct CASE WHEN thankyou_made_it = 1 THEN website_session_id ELSE NULL END)/count(distinct CASE WHEN billing_made_it = 1 THEN website_session_id ELSE NULL END) AS  billing_click_rt
from session_level_made_it_flags_assignment2;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

sessions	products_count	mrfuzzy_count	cart_count	shipping_count	billing_count	thankyou_count
282706	155823	97413	56148	37961	2621	18822