

## Project Work - Web Analytics

The project uses MySQL to perform analysis. The code for database connection is hidden for security reasons.

```
In [1]: %load_ext sql
import mysql
```

```
In [2]: # Connecting to the database
%sql mysql://root: @localhost/mavenfuzzyfactory
```

### Analysing Traffic Sources

Traffic source analysis is done with the help of **utm** tracker information.

```
In [3]: # utm tracker information can be viewed for a given website session.
```

```
In [4]: %%sql

SELECT *
FROM website_sessions
WHERE website_session_id = 1059;

* mysql://root:***@localhost/mavenfuzzyfactory
1 rows affected.
```

```
Out[4]:
```

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

```
In [5]: # All available traffic sources in the database.
```

In [6]: %%sql

```
SELECT DISTINCT utm_source, utm_campaign, utm_content
FROM website_sessions;
```

```
* mysql://root:***@localhost/mavenfuzzyfactory
7 rows affected.
```

Out[6]:

utm_source	utm_campaign	utm_content
------------	--------------	-------------

gsearch	nonbrand	g_ad_1
None	None	None
gsearch	brand	g_ad_2
bsearch	brand	b_ad_2
bsearch	nonbrand	b_ad_1
socialbook	pilot	social_ad_1
socialbook	desktop_targeted	social_ad_2

Where is the bulk of traffic for the business coming from?

- *The query is asked on April 12, 2012*

In [7]: %%sql

```
SELECT utm_source,
       utm_campaign,
       utm_content,
       http_referer,
       COUNT(utm_source) AS 'Sessions'
FROM website_sessions
WHERE created_at <= '2012-04-12' AND utm_source IS NOT NULL
GROUP BY utm_source, utm_campaign, utm_content, http_referer;
```

```
* mysql://root:***@localhost/mavenfuzzyfactory
3 rows affected.
```

Out[7]:

utm_source	utm_campaign	utm_content	http_referer	Sessions
------------	--------------	-------------	--------------	----------

gsearch	nonbrand	g_ad_1	https://www.gsearch.com	3613
gsearch	brand	g_ad_2	https://www.gsearch.com	26
bsearch	brand	b_ad_2	https://www.bsearch.com	7

In [8]: *# It is found that most of the traffic is coming from 'gsearch nonbrand' campaign.*

Is 'gsearch nonbrand' generating sales? What is the **conversion rate**?

- The query is asked on April 14, 2012

- The minimum CVR threshold is 4% (to justify budget)

In [9]: `%%sql`

```
SELECT
    order_id,
    o.created_at,
    o.website_session_id,
    ws.utm_source,
    ws.utm_campaign,
    ws.utm_content
FROM orders o
LEFT JOIN website_sessions ws
ON o.website_session_id = ws.website_session_id
WHERE
    ws.utm_source = 'gsearch' AND
    ws.utm_campaign = 'nonbrand' AND
    ws.created_at <= '2012-04-14'
LIMIT 5
;
```

\* mysql://root:\*\*\*@localhost/mavenfuzzyfactory

5 rows affected.

Out[9]:

	order_id	created_at	website_session_id	utm_source	utm_campaign	utm_content
1	2012-03-19 10:42:46	20	gsearch	nonbrand	g_ad_1	
2	2012-03-19 19:27:37	104	gsearch	nonbrand	g_ad_1	
3	2012-03-20 06:44:45	147	gsearch	nonbrand	g_ad_1	
4	2012-03-20 09:41:45	160	gsearch	nonbrand	g_ad_1	
5	2012-03-20 11:28:15	177	gsearch	nonbrand	g_ad_1	

In [10]: %%sql

```
SELECT
    COUNT(DISTINCT ws.website_session_id) AS 'Sessions',
    COUNT(DISTINCT o.order_id) AS 'Orders',
    (COUNT(DISTINCT o.order_id)/COUNT(DISTINCT ws.website_session_id))*100 AS 'Conversion Rate %'
FROM website_sessions ws
LEFT JOIN orders o
ON ws.website_session_id = o.website_session_id
WHERE
    ws.utm_source = 'gsearch' AND
    ws.utm_campaign = 'nonbrand' AND
    ws.created_at <= '2012-04-14';
```

```
* mysql://root:***@localhost/mavenfuzzyfactory
1 rows affected.
```

Out[10]:

Sessions	Orders	Conversion Rate %
3895	112	2.8755

In [11]: *# It is observed that CVR is below the 4% threshold.*  
*# Hence, the money is overspent and search bids must be reduced.*

Based on the analysis, **gsearch nonbrand** was *bid down* on 15 April, 2012

- Have bid changes caused a change in volume?
- The query is asked on 10 May, 2012

In [12]: %%sql

```
SELECT
    MIN DATE(created_at) AS 'Week Start Date',
    COUNT(website_session_id) AS 'Sessions'
FROM website_sessions ws
WHERE
    ws.utm_source = 'gsearch' AND
    ws.utm_campaign = 'nonbrand' AND
    ws.created_at <= '2012-05-10'
GROUP BY WEEK(created_at);
```

```
* mysql://root:***@localhost/mavenfuzzyfactory
8 rows affected.
```

Out[12]:

Week Start Date	Sessions
-----------------	----------

2012-03-19	896
2012-03-25	956
2012-04-01	1152
2012-04-08	983
2012-04-15	621
2012-04-22	594
2012-04-29	681
2012-05-06	399

In [13]: # 'gsearch nonbrand' traffic is sensitive to bid  
# changes since session volume has slightly dropped after Apr 15

Could **device type** influence conversion rates?

- Retrieve CVR by device type
- The query is asked on 11 May, 2012

In [14]: %%sql

```
SELECT
    ws.device_type,
    COUNT(DISTINCT ws.website_session_id) AS 'Sessions',
    COUNT(DISTINCT o.order_id) AS 'Orders',
    (COUNT(DISTINCT o.order_id)/COUNT(DISTINCT ws.website_session_id))*100 AS 'Conversion Rate %'
FROM website_sessions ws
LEFT JOIN orders o
ON ws.website_session_id = o.website_session_id
WHERE
    ws.utm_source = 'gsearch' AND
    ws.utm_campaign = 'nonbrand' AND
    ws.created_at <= '2012-05-11'
GROUP BY 1;
```

```
* mysql://root:***@localhost/mavenfuzzyfactory
2 rows affected.
```

Out[14]:

device_type	Sessions	Orders	Conversion Rate %
desktop	3911	146	3.7331
mobile	2492	24	0.9631

In [15]: *# Yes, device type influences CVR. It's best to bid-up only for desktop.*

The bid on **gsearch nonbrand desktop** campaigns was increased on 19 May, 2012.

- *What impact is observed on sessions volume?*
- *The request was made on 09 June, 2012*

In [16]: %%sql

```
SELECT
    MIN(DATE(created_at)) AS 'Week Start Date',
    COUNT(CASE WHEN device_type = 'desktop' THEN website_session_id ELSE NULL END) AS 'Desktop Sessions',
    COUNT(CASE WHEN device_type = 'mobile' THEN website_session_id ELSE NULL END) AS 'Mobile Sessions'
FROM website_sessions ws
WHERE
    ws.utm_source = 'gsearch' AND
    ws.utm_campaign = 'nonbrand' AND
    ws.created_at BETWEEN '2012-04-15' AND '2012-06-09'
GROUP BY WEEK(created_at);
```

```
* mysql://root:***@localhost/mavenfuzzyfactory
8 rows affected.
```

Out[16]:

Week Start Date	Desktop Sessions	Mobile Sessions
2012-04-15	383	238
2012-04-22	360	234
2012-04-29	425	256
2012-05-06	430	282
2012-05-13	403	214
2012-05-20	661	190
2012-05-27	585	183
2012-06-03	582	157

In [19]:

```
# 'gsearch nonbrand desktop' was bid-up on 19 Apr, 2012.
# Desktop session volume has increased overall after bidding up.
```