# E-COMMERCE & WEB ANALYTICS USING MySQL

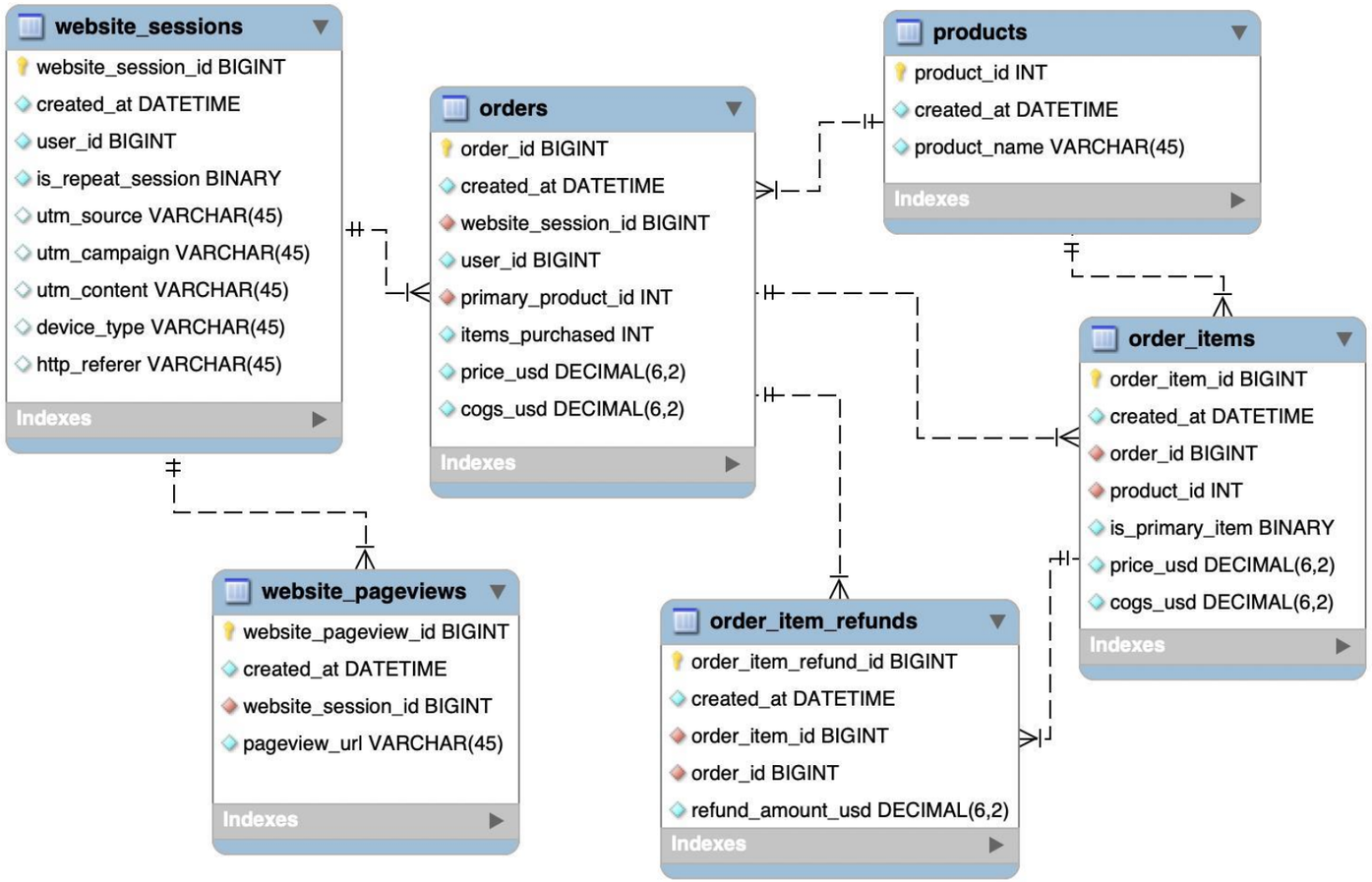
Md Tauhid Alam

# BUSINESS SCENARIO

As a newly hired e-commerce database analyst for Maven Fuzzy Factory;

- analyze and optimize marketing channels,
- measure and test website conversion performance,
- and use data to understand the impact of new product launches.

# GETTING TO KNOW THE DATABASE




*Data Source: Excel Maven & Maven Analytics, LLC*

# ANALYZING WEBSITE TRAFFIC SOURCES

In the very first section, we'll be analyzing the search data (**utm_parameters**) to understand user behavior patterns and take budget shifting decisions to **optimize marketing channel performance**.

NOTE:

Paid traffic is commonly tagged with tracking (UTM) parameters, which are appended to URLs and allow us to tie website activity back to specific traffic sources and campaigns

www.abcwebsite.com?utm_source=trafficSource&utm_campaign=campaignName

**Session-to-Order-Conversion Rate (CVR)** is the ratio of number of distinct orders to number of distinct website sessions.

**NEW MESSAGE**
*April 12, 2012*

From: **Cindy Sharp** *(CEO)*
Subject: **Site traffic breakdown**

Good morning,

We've been live for almost a month now and we're starting to generate sales. Can you help me understand where the bulk of our website sessions are coming from, through yesterday?

I'd like to see a breakdown by **UTM source**, **campaign** and **referring domain** if possible. Thanks!

-Cindy

Reply Forward

```sql
SELECT
    ws.utm_source,
    ws.utm_campaign,
    ws.http_referer AS referring_domain,
    COUNT( DISTINCT ws.website_session_id) AS num_sessions
FROM website_sessions ws
WHERE ws.created_at < '2012-04-12'
GROUP BY
    ws.utm_source,
    ws.utm_campaign,
    ws.http_referer
ORDER BY num_sessions DESC;
```

| utm_source | utm_campaign | referring_domain | num_sessions |
|---|---|---|---|
| gsearch | nonbrand | https://www.gsearch.com | 3613 |
| NULL | NULL | NULL | 28 |
| NULL | NULL | https://www.gsearch.com | 27 |
| gsearch | brand | https://www.gsearch.com | 26 |
| NULL | NULL | https://www.bsearch.com | 7 |
| bsearch | brand | https://www.bsearch.com | 7 |

*Drill deeper into **gsearch nonbrand** campaign traffic to explore potential optimization opportunities.*

**NEW MESSAGE**
*April 14, 2012*

From: **Tom Parmesan** *(Marketing Director)*
Subject: **Gsearch conversion rate**

Hi there,

Sounds like gsearch nonbrand is our major traffic source, but we need to understand if those sessions are driving sales.

Could you please **calculate the conversion rate (CVR) from session to order**? Based on what we're paying for clicks, we'll need a CVR of **at least 4%** to make the numbers work.

If we're much lower, we'll need to reduce bids. If we're higher, we can increase bids to drive more volume.

Thanks, Tom

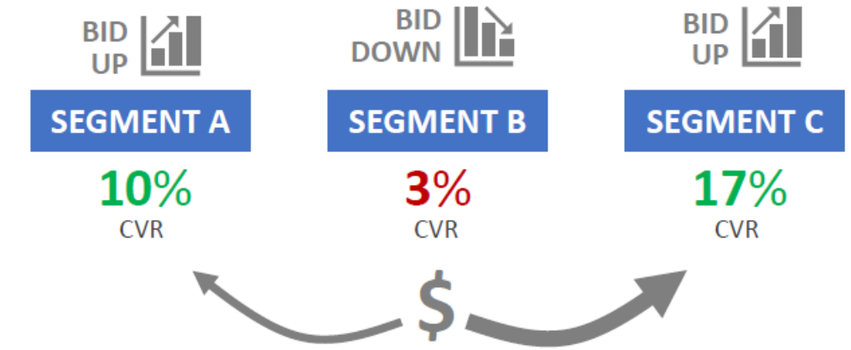
Reply Forward

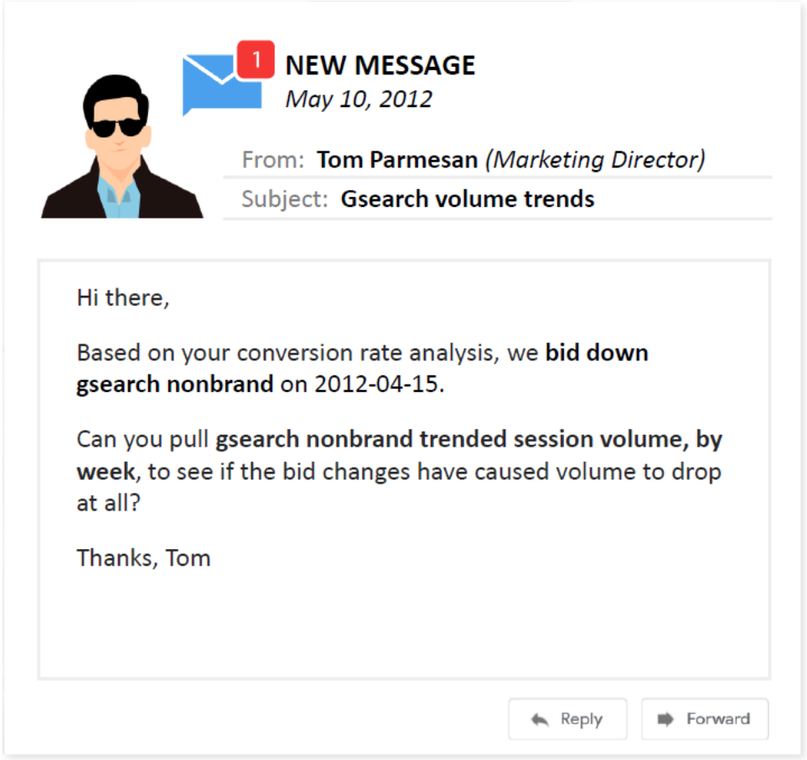
```sql
SELECT
    COUNT( DISTINCT ws.website_session_id) AS num_sessions,
    COUNT( DISTINCT o.order_id) AS num_orders,
    ROUND( COUNT(DISTINCT o.order_id)/COUNT(DISTINCT ws.website_session_id)*100, 2 ) AS conversion_rate
FROM website_sessions ws
    LEFT JOIN orders o ON ws.website_session_id = o.website_session_id
WHERE
    ws.created_at < '2012-04-14'
    AND ws.utm_source = 'gsearch'
    AND ws.utm_campaign = 'nonbrand';
```

| num_sessions | num_orders | conversion_rate |
|---|---|---|
| 3895 | 112 | 2.88 |

**CVR** less than 4% threshold implies that we're over-spending on our search bids, and thus need to reduce it.

*Monitor the impact of bid reductions and **analyze performance trending by device type** in order to refine bidding strategy.*





**NEW MESSAGE**
*May 10, 2012*

From: **Tom Parmesan** *(Marketing Director)*
Subject: **Gsearch volume trends**

Hi there,

Based on your conversion rate analysis, we **bid down gsearch nonbrand** on 2012-04-15.

Can you pull **gsearch nonbrand trended session volume, by week**, to see if the bid changes have caused volume to drop at all?

Thanks, Tom

Reply Forward

```sql
SELECT
    MIN( DATE(created_at) ) AS start_of_week,
    WEEK(created_at) AS week_num,
    COUNT(website_session_id) AS num_sessions
FROM website_sessions
WHERE
    created_at < '2012-5-10' AND
    utm_source = 'gsearch' AND utm_campaign = 'nonbrand'
GROUP BY
    YEAR(created_at), WEEK(created_at)
ORDER BY 2 DESC;
```

| start_of_week | week_num | num_sessions |
|---|---|---|
| 2012-05-06 | 19 | 399 |
| 2012-04-29 | 18 | 681 |
| 2012-04-22 | 17 | 594 |
| 2012-04-15 | 16 | 621 |
| 2012-04-08 | 15 | 983 |
| 2012-04-01 | 14 | 1152 |
| 2012-03-25 | 13 | 956 |
| 2012-03-19 | 12 | 896 |

session volume for **gsearch nonbrand** is fairly sensitive to bid changes.

*Since we want maximum session volume without over-spending on ads, we should **continue to monitor volume levels** and think about additional ways to make our campaigns more efficient.*

**NEW MESSAGE**
*May 11, 2012*

From: **Tom Parmesan** *(Marketing Director)*
Subject: **Gsearch device-level performance**

Hi there,

I was trying to use our site on my mobile device the other day, and the experience was not great.

Could you pull **conversion rates from session to order**, by **device type**?

If desktop performance is better than on mobile we may be able to bid up for desktop specifically to get more volume?

Thanks, Tom

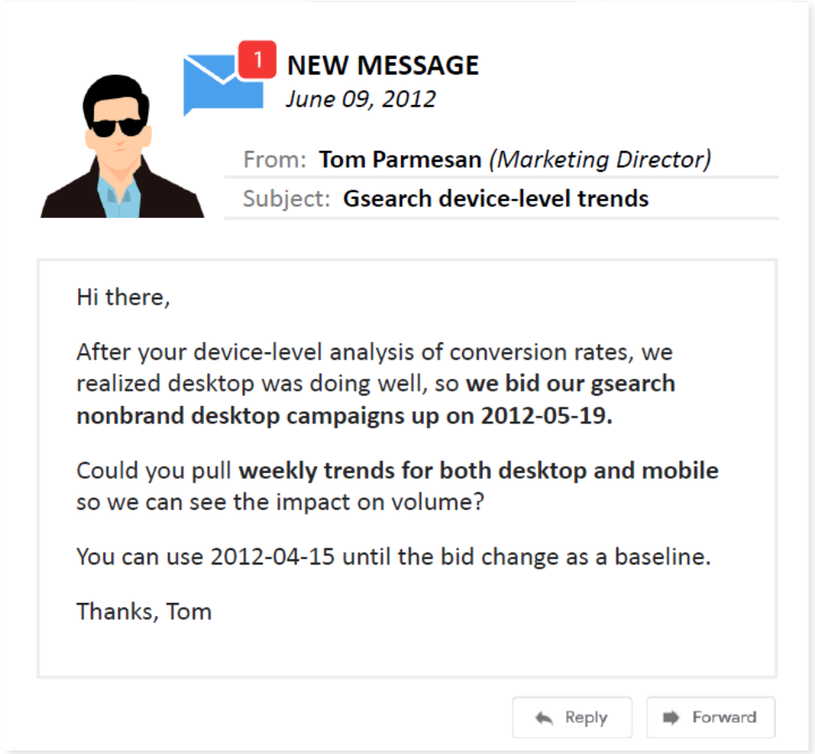
Reply Forward

```sql
SELECT
    ws.device_type,
    COUNT(DISTINCT ws.website_session_id) AS sessions,
    COUNT(DISTINCT o.order_id) AS orders,
    ROUND( COUNT(DISTINCT o.order_id)/COUNT(DISTINCT ws.website_session_id)*100, 2 ) AS CVR_pct
FROM website_sessions ws
    LEFT JOIN orders o ON ws.website_session_id = o.website_session_id
WHERE
    ws.created_at < '2012-05-11' AND
    ws.utm_source = 'gsearch' AND utm_campaign = 'nonbrand'
GROUP BY 1
ORDER BY 2 DESC;
```

| device_type | sessions | orders | CVR_pct |
|---|---|---|---|
| desktop | 3911 | 146 | 3.73 |
| mobile | 2492 | 24 | 0.96 |

***Increase bids on desktop*** *and analyze volume by device type to* ***see if the bid changes make a material impact*** *while continuously looking optimization opportunities.*

traffic conversion rate for **desktop sessions** is very **close to the threshold of 4%** while it is much lower for mobile devices



**NEW MESSAGE**
*June 09, 2012*

From: **Tom Parmesan** *(Marketing Director)*
Subject: **Gsearch device-level trends**

Hi there,

After your device-level analysis of conversion rates, we realized desktop was doing well, so **we bid our gsearch nonbrand desktop campaigns up on 2012-05-19.**

Could you pull **weekly trends for both desktop and mobile** so we can see the impact on volume?

You can use 2012-04-15 until the bid change as a baseline.

Thanks, Tom

Reply Forward

```sql
SELECT
    MIN(DATE(ws.created_at)) AS start_of_wk,
    WEEK(ws.created_at) AS wk_of_yr,
    COUNT(DISTINCT CASE WHEN ws.device_type = 'desktop' THEN ws.website_session_id ELSE NULL END) AS desktop_sessions,
    COUNT(DISTINCT CASE WHEN ws.device_type = 'mobile' THEN ws.website_session_id ELSE NULL END) AS mobile_sessions
FROM website_sessions ws
WHERE
    ws.created_at < '2012-06-09'
    AND ws.created_at > '2012-04-15'
    AND ws.utm_source = 'gsearch'
    AND ws.utm_campaign = 'nonbrand'
GROUP BY
    YEAR(ws.created_at),
    WEEK(ws.created_at)
ORDER BY 1 DESC;
```

| start_of_wk | wk_of_yr | desktop_sessions | mobile_sessions |
|---|---|---|---|
| 2012-06-03 | 23 | 582 | 157 |
| 2012-05-27 | 22 | 585 | 183 |
| 2012-05-20 | 21 | 661 | 190 |
| 2012-05-13 | 20 | 403 | 214 |
| 2012-05-06 | 19 | 430 | 282 |
| 2012-04-29 | 18 | 425 | 256 |
| 2012-04-22 | 17 | 360 | 234 |
| 2012-04-15 | 16 | 383 | 238 |

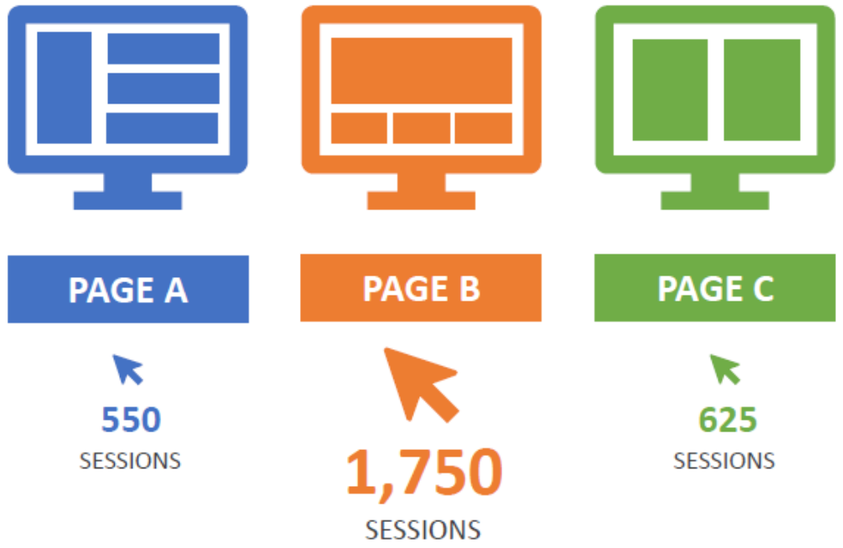
**significant increase** in desktop session volumes as compared to previous levels

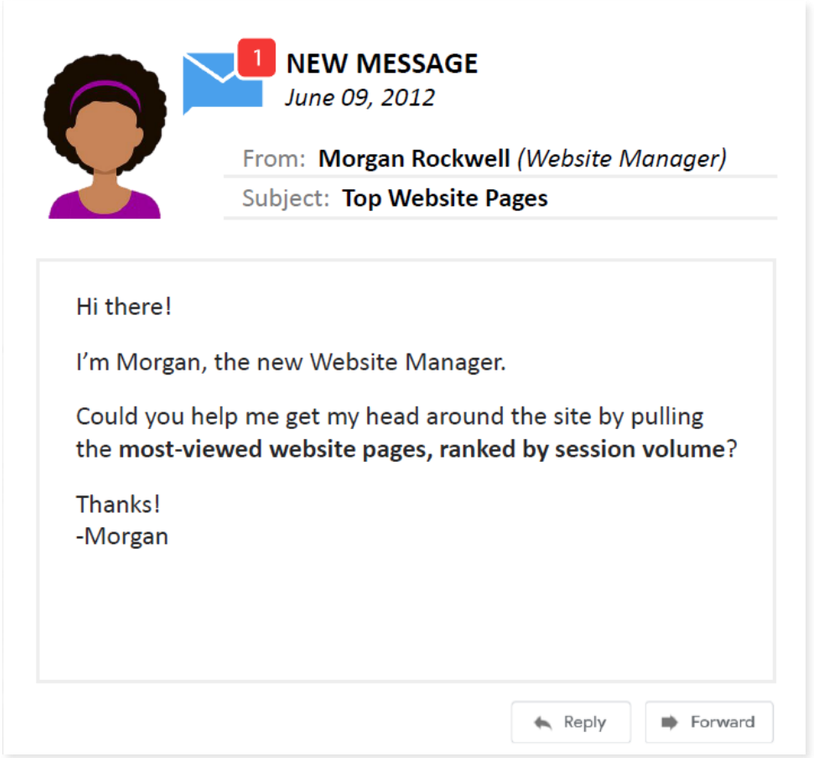
*The bid optimization led to effective results. Keep monitoring the **impact of bid changes over time on device level session volume and conversion rates** to further optimize spend.*

# ANALYZING WEBSITE PERFORMANCE

In this section, we'll be analyzing the performance of web-pages and landing (or entry pages) to understand and optimize user experience via bounce rate and conversion funnel analysis.

We'll also be conducting A/B testing with new website landing pages to see if the performance can be improved.

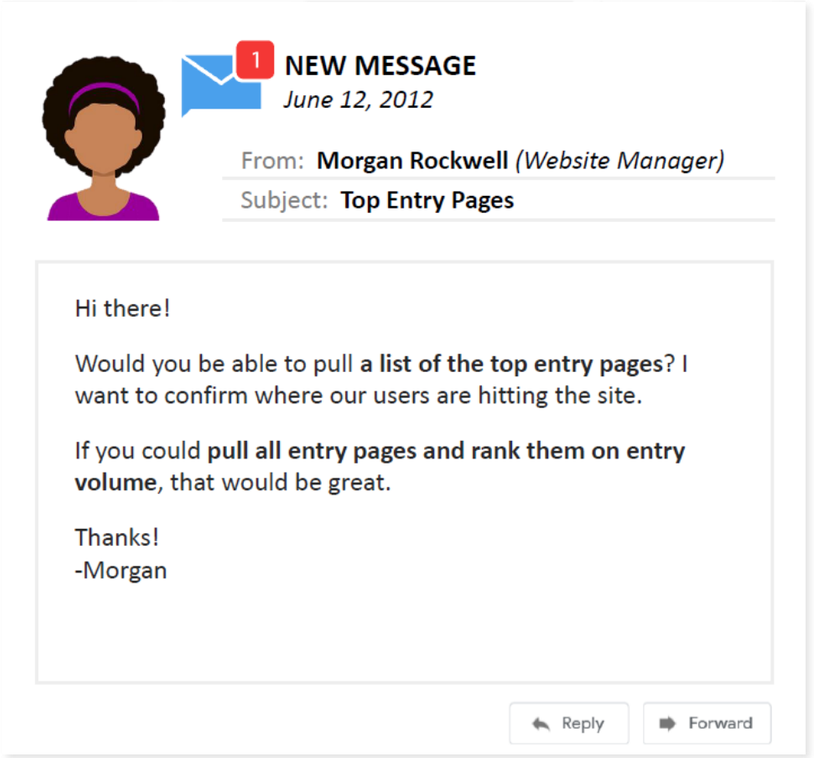





```sql
-- most viewed web-pages ranked by session volume
SELECT
    wp.pageview_url AS web_pages,
    COUNT(DISTINCT wp.website_session_id) AS num_sessions
FROM
    website_pageviews wp
WHERE
    wp.created_at < '2012-06-09'
GROUP BY 1
ORDER BY 2 DESC;
```

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

*Dig into whether this list is also representative of our **top entry pages** and analyze the performance of **each of our top pages** to look for **improvement opportunities***




```sql
CREATE TEMPORARY TABLE first_pageviews  -- contains first pageview for each session
SELECT
    wp.website_session_id,
    MIN(wp.website_pageview_id) AS first_pv_id
FROM website_pageviews wp
WHERE wp.created_at < '2012-06-12'
GROUP BY wp.website_session_id;
```

| website_session_id | first_pv_id |
|---|---|
| 64 | 113 |
| 65 | 116 |
| 66 | 117 |
| 67 | 118 |
| 68 | 119 |
| 69 | 120 |

```sql
-- returns the first pageview_url(s) along with the sessions_hitting_that_url
SELECT
    wp.pageview_url AS entry_pages,
    COUNT(DISTINCT fp.website_session_id) AS entry_volume
FROM
    first_pageviews fp
    LEFT JOIN website_pageviews wp ON fp.first_pv_id = wp.website_pageview_id
GROUP BY 1
ORDER BY 2 DESC;
```

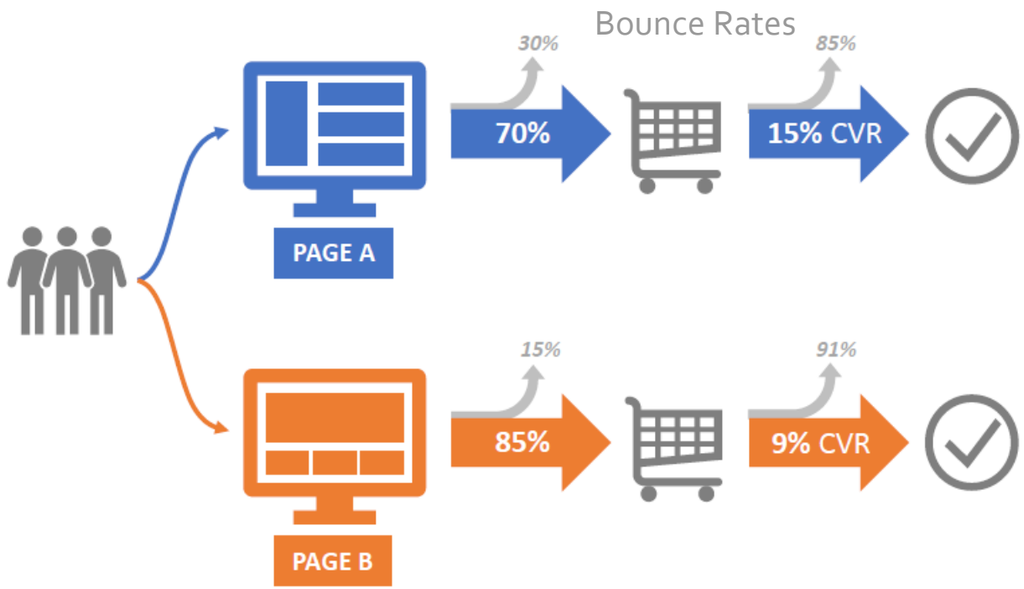
| entry_pages | entry_volume |
|---|---|
| /home | 10714 |

all traffic lands only on the **home** page first

*Analyze landing page performance (for the **homepage** specifically) and think about whether or not the homepage is the **best initial experience** for all customers.*

# **BUSINESS CONCEPT:** LANDING PAGE PERFORMANCE AND TESTING

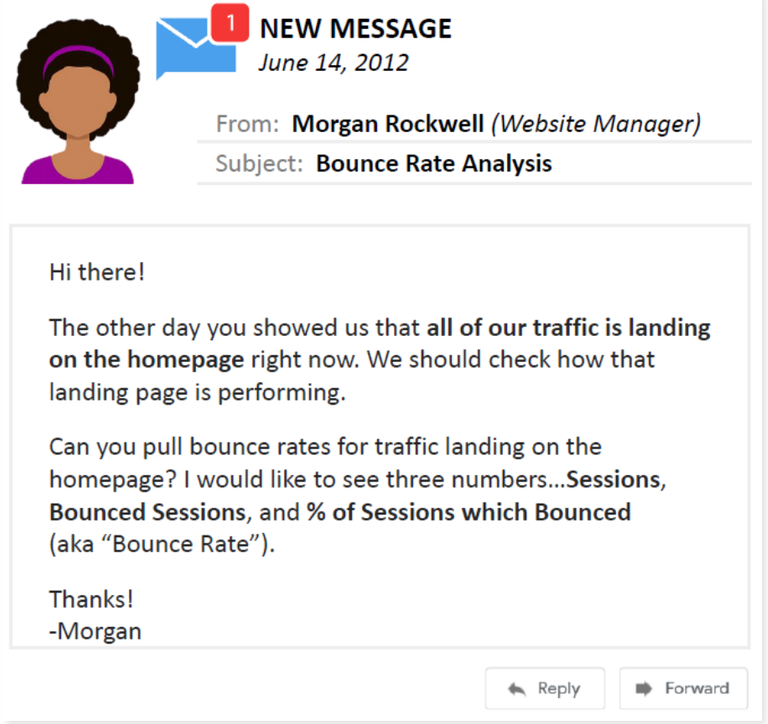
Landing page analysis and testing is about understanding the performance of your key landing pages and then testing to improve your results




*USE CASES:*

- *Identifying improvement opportunities for less performing pages*
- *A/B analysis for deciding the right version of landing page*

We will be using **temporary tables** again to write multi-step queries for landing page performance analysis.



**NEW MESSAGE**
*June 14, 2012*

From: **Morgan Rockwell** *(Website Manager)*
Subject: **Bounce Rate Analysis**

Hi there!

The other day you showed us that **all of our traffic is landing on the homepage** right now. We should check how that landing page is performing.

Can you pull bounce rates for traffic landing on the homepage? I would like to see three numbers...**Sessions**, **Bounced Sessions**, and **% of Sessions which Bounced** (aka "Bounce Rate").

Thanks!
-Morgan

Reply Forward

```sql
-- Step 1: Finding first_pv_ids for each relevant session
CREATE TEMPORARY TABLE first_pageviews
SELECT
    ws.website_session_id,
    MIN(wp.website_pageview_id) AS first_pv_id
FROM
    website_sessions ws
    LEFT JOIN website_pageviews wp ON ws.website_session_id = wp.website_session_id
WHERE
    ws.created_at < '2012-06-14'
GROUP BY
    1;
```

| website_session_id | first_pv_id |
|---|---|
| 34 | 55 |
| 35 | 56 |
| 36 | 59 |
| 37 | 60 |
| 38 | 61 |

11,048 sessions

↓

All have **/home** as the landing page

```sql
-- Step 2: Identifying sessions with landing pages by linking to urls
CREATE TEMPORARY TABLE sessions_w_home_lp
SELECT
    fp.website_session_id,
    wp.pageview_url AS landing_pgs
FROM
    first_pageviews fp
    LEFT JOIN website_pageviews wp ON fp.first_pv_id = wp.website_pageview_id
WHERE
    wp.pageview_url = '/home';
```

| website_session_id | landing_pgs |
|---|---|
| 1 | /home |
| 2 | /home |
| 3 | /home |
| 4 | /home |
| 5 | /home |

↓

6,538 bounced sessions

```sql
-- Step 3: Counting pageviews for each session to identify bounced_sessions
CREATE TEMPORARY TABLE bounced_sessions_only
SELECT
    sl.website_session_id,
    sl.landing_pgs,
    COUNT(wp.website_pageview_id) AS num_pgs_viewed
FROM
    sessions_w_home_lp sl
    LEFT JOIN website_pageviews wp ON sl.website_session_id = wp.website_session_id
GROUP BY
    1, 2
HAVING
    num_pgs_viewed = 1; -- limiting to bounced sessions only
```

| website_session_id | landing_pgs | num_pgs_viewed |
|---|---|---|
| 1 | /home | 1 |
| 2 | /home | 1 |
| 3 | /home | 1 |
| 4 | /home | 1 |
| 5 | /home | 1 |

```sql
-- Step 4: Summarizing Results
SELECT
    COUNT(DISTINCT sl.website_session_id) AS total_sessions,
    COUNT(DISTINCT bs.website_session_id) AS bounced_sessions,
    COUNT(DISTINCT bs.website_session_id)/COUNT(DISTINCT sl.website_session_id) AS bounce_rate
FROM
    sessions_w_home_lp sl
    LEFT JOIN bounced_sessions_only bs ON sl.website_session_id = bs.website_session_id;
```

| total_sessions | bounced_sessions | bounce_rate |
|---|---|---|
| 11048 | 6538 | 0.5918 |

**60%** bounce rate is **pretty high for paid traffic**, conduct **experiment** with a **custom landing page** to see if we can improve results.

**NEW MESSAGE**

*July 28, 2012*

From: **Morgan Rockwell** *(Website Manager)*

Subject: **Help Analyzing LP Test**

Hi there!

Based on your bounce rate analysis, we ran a new custom landing page (**/lander-1**) in a **50/50 test against the homepage** (**/home**) for our gsearch nonbrand traffic.

Can you **pull bounce rates for the two groups** so we can evaluate the new page? Make sure to **just look at the time period where /lander-1 was getting traffic**, so that it is a fair comparison.

Thanks, Morgan

```sql
-- Step 1: Finding the first_pv_ids for each relevant sessions
CREATE TEMPORARY TABLE first_test_pageviews
SELECT
    ws.website_session_id,
    MIN(wp.website_pageview_id) AS first_pv_id
FROM
    website_sessions ws
    LEFT JOIN website_pageviews wp ON ws.website_session_id = wp.website_session_id
WHERE
    ws.created_at < '2014-07-28'
    AND wp.website_pageview_id > (
        SELECT
            MIN(wp.website_pageview_id)     -- This subquery yields 23504 as the result,
        FROM
            website_pageviews wp            -- which is the first_pv_id when '/lander-1'
        WHERE
            wp.pageview_url = '/lander-1'   -- appeared for the first time and this first_pv_id
            AND wp.created_at IS NOT NULL
    )                                       -- is used to set up the correct analysis timeframe.
    AND ws.utm_source = 'gsearch'
    AND ws.utm_campaign = 'nonbrand'
GROUP BY
    1;
```

| website_session_id | first_pv_id |
|---|---|
| 11684 | 23505 |
| 11685 | 23506 |
| 11686 | 23507 |
| 11687 | 23509 |
| 11688 | 23510 |

Total **172,646** sessions in the timeframe

```sql
-- Step 2: Identifying the landing_page_urls
-- by linking first_pv_ids for each session
CREATE TEMPORARY TABLE test_sessions_with_lp
SELECT
    fp.website_session_id,
    wp.pageview_url AS landing_pgs
FROM
    first_test_pageviews fp
    LEFT JOIN website_pageviews wp
    ON fp.first_pv_id = wp.website_pageview_id
WHERE
    wp.pageview_url IN ('/home','/lander-1');
```

| website_session_id | landing_pgs |
|---|---|
| 11684 | /home |
| 11685 | /lander-1 |
| 11686 | /lander-1 |
| 11687 | /home |
| 11688 | /home |

**40,442** records

```sql
-- Step 3: Counting pageviews for each session to identify bounced_sessions
CREATE TEMPORARY TABLE test_bounced_sessions
SELECT
    sl.website_session_id,
    sl.landing_pgs,
    COUNT(wp.website_pageview_id) AS num_pgs_viewed
FROM
    test_sessions_with_lp sl
    LEFT JOIN website_pageviews wp
    ON sl.website_session_id = wp.website_session_id
GROUP BY
    1, 2
HAVING
    num_pgs_viewed = 1; -- limiting to just "bounced" sessions
```

| website_session_id | landing_pgs | num_pgs_viewed |
|---|---|---|
| 11684 | /home | 1 |
| 11685 | /lander-1 | 1 |
| 11687 | /home | 1 |
| 11688 | /home | 1 |

**21,411** records

```sql
-- Step 4: Summarizing Results
SELECT
    sl.landing_pgs,
    COUNT(DISTINCT sl.website_session_id) as total_sessions,
    COUNT(DISTINCT bs.website_session_id) as bounced_sessions,
    COUNT(DISTINCT bs.website_session_id) /
        COUNT(DISTINCT sl.website_session_id) AS bounce_rates
FROM
    test_sessions_with_lp sl
    LEFT JOIN test_bounced_sessions bs
    ON sl.website_session_id = bs.website_session_id
GROUP BY
    1;
```

| landing_pgs | total_sessions | bounced_sessions | bounce_rates |
|---|---|---|---|
| /home | 2328 | 1365 | 0.5863 |
| /lander-1 | 38114 | 20046 | 0.5259 |

The new '**/lander-1**' has around **6% less bounce rate**, and thus can be considered an option for the **nonbrand** paid traffic campaign.

**NEW MESSAGE**
*August 31, 2012*

From: **Morgan Rockwell** *(Website Manager)*
Subject: **Landing Page Trend Analysis**

Hi there,

Could you pull the volume of **paid search nonbrand traffic landing on /home and /lander-1, trended weekly since June 1st?** I want to confirm the traffic is all routed correctly.

Could you also **pull our overall paid search bounce rate trended weekly**? I want to make sure the lander change has improved the overall picture.

Thanks!

```sql
-- # Step 1: Get first_pv_ids and num_pgs_viewed for each relevant session
CREATE TEMPORARY TABLE sessions_w_fp_and_npv
SELECT
    ws.website_session_id,
    MIN(wp.website_pageview_id) AS first_pv_id,
    COUNT(wp.website_pageview_id) AS num_pgs_viewed
FROM
    website_sessions ws
LEFT JOIN
    website_pageviews wp ON ws.website_session_id = wp.website_session_id
WHERE
    ws.created_at BETWEEN '2012-06-01' AND '2012-08-31'
    AND ws.utm_source = 'gsearch'
    AND ws.utm_campaign = 'nonbrand'
GROUP BY
    1; -- 11624 rows
```

| website_session_id | first_pv_id | num_pgs_viewed |
|---|---|---|
| 9350 | 18598 | 3 |
| 9351 | 18600 | 3 |
| 9352 | 18601 | 4 |
| 9354 | 18611 | 1 |
| 9356 | 18616 | 6 |

11,624 records

```sql
-- # Step 2: Identifying landing pages for each session on each date in the timeframe
CREATE TEMPORARY TABLE sessions_w_lp_and_ca
SELECT
    wp.created_at,
    fp.website_session_id,
    wp.pageview_url
FROM
    sessions_w_fp_and_npv fp
LEFT JOIN
    website_pageviews wp ON fp.first_pv_id = wp.website_pageview_id;
```

| created_at | website_session_id | pageview_url |
|---|---|---|
| 2012-07-13 22:50:14 | 15050 | /home |
| 2012-07-13 23:44:06 | 15052 | /lander-1 |
| 2012-07-14 01:48:49 | 15053 | /lander-1 |
| 2012-07-14 02:05:33 | 15054 | /home |
| 2012-07-14 02:08:33 | 15055 | /lander-1 |

```sql
-- # Step 3: Summarizing results grouped by wk_of_yr to show weekly trend
SELECT
    WEEK(sl.created_at) AS wk_of_yr,
    MIN(DATE(sl.created_at)) AS st_of_wk,
    COUNT(DISTINCT CASE WHEN sl.pageview_url = '/home' THEN fp.website_session_id ELSE NULL END) AS home_sessions,
    COUNT(DISTINCT CASE WHEN sl.pageview_url = '/lander-1' THEN fp.website_session_id ELSE NULL END) AS lander_1_sessions,
    COUNT(DISTINCT fp.website_session_id) AS total_sessions,
    COUNT(DISTINCT CASE WHEN fp.num_pgs_viewed = 1 THEN fp.website_session_id ELSE NULL END) AS bounced_sessions,
    ROUND( COUNT(DISTINCT CASE WHEN fp.num_pgs_viewed = 1 THEN fp.website_session_id ELSE NULL END) /
        COUNT(DISTINCT fp.website_session_id) * 100, 2 )AS overall_bounce_rate_pct
FROM
    sessions_w_fp_and_npv fp
LEFT JOIN
    sessions_w_lp_and_ca sl ON fp.website_session_id = sl.website_session_id
GROUP BY
    1
ORDER BY
    1 DESC;
```

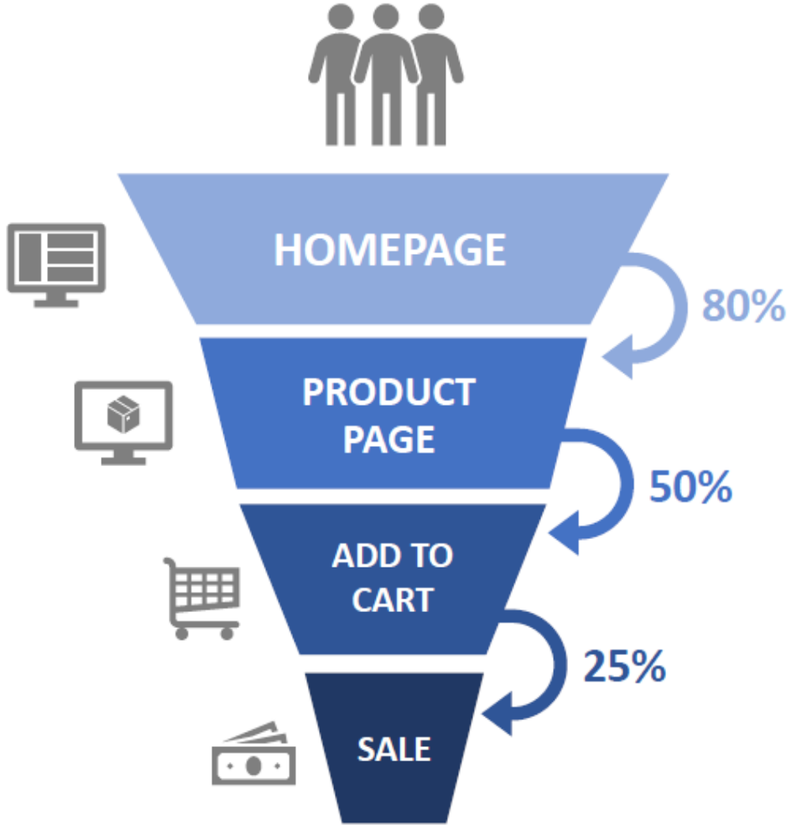
| wk_of_yr | st_of_wk | home_sessions | lander_1_sessions | total_sessions | bounced_sessions | overall_bounce_rate_pct |
|---|---|---|---|---|---|---|
| 35 | 2012-08-26 | 0 | 833 | 833 | 448 | 53.78 |
| 34 | 2012-08-19 | 0 | 1012 | 1012 | 507 | 50.10 |
| 33 | 2012-08-12 | 0 | 998 | 998 | 513 | 51.40 |
| 32 | 2012-08-05 | 0 | 1087 | 1087 | 585 | 53.82 |
| 31 | 2012-07-29 | 33 | 995 | 1028 | 511 | 49.71 |
| 30 | 2012-07-22 | 402 | 394 | 796 | 409 | 51.38 |
| 29 | 2012-07-15 | 429 | 421 | 850 | 461 | 54.24 |
| 28 | 2012-07-08 | 390 | 411 | 801 | 454 | 56.68 |
| 27 | 2012-07-01 | 392 | 388 | 780 | 454 | 58.21 |
| 26 | 2012-06-24 | 369 | 386 | 755 | 440 | 58.28 |
| 25 | 2012-06-17 | 492 | 350 | 842 | 470 | 55.82 |
| 24 | 2012-06-10 | 875 | 0 | 875 | 539 | 61.60 |
| 23 | 2012-06-03 | 792 | 0 | 792 | 465 | 58.71 |
| 22 | 2012-06-01 | 175 | 0 | 175 | 106 | 60.57 |

Both pages were getting traffic for a while

*The implementation of custom lander was a success as overall **bounce rate has improved over time** by around **7%**.*

# **BUSINESS CONCEPT:** ANALYZING & TESTING CONVERSION FUNNELS

Conversion funnel analysis is about **understanding and optimizing each step of your user's experience on their journey toward purchasing your products**




**NEW MESSAGE**
*September 05, 2012*

From: **Morgan Rockwell** *(Website Manager)*
Subject: **Help Analyzing Conversion Funnels**

Hi there!

I'd like to understand where we lose our gsearch visitors between the new /lander-1 page and placing an order. Can you **build us a full conversion funnel, analyzing how many customers make it to each step**?

Start with **/lander-1** and build the funnel all the way to our **thank you page**. Please use data since **August 5th**.

Thanks!
-Morgan

```sql
-- # Step 1: Identifying relevant sessions and bringing in relevant pv_ids
CREATE TEMPORARY TABLE session_level_flags
SELECT
    website_session_id,
    MAX(products_pg) AS made_to_products,    -- When we use GROUP BY we need
    MAX(fuzzy_pg) AS made_to_fuzzy,          -- to use aggregate functions in our select statement
    MAX(cart_pg) AS made_to_cart,            -- for any columns not named in the GROUP BY.
    MAX(shipping_pg) AS made_to_shipping,    -- That's why we are using MAX()
    MAX(billing_pg) AS made_to_billing,
    MAX(thank_you_pg) AS made_to_thank_you
FROM
    (
        SELECT
            ws.website_session_id,
            wp.created_at AS pv_created_at,
            wp.pageview_url,
            CASE WHEN wp.pageview_url = '/products' THEN 1 ELSE 0 END AS products_pg,
            CASE WHEN wp.pageview_url = '/the-original-mr-fuzzy' THEN 1 ELSE 0 END AS fuzzy_pg,
            CASE WHEN wp.pageview_url = '/cart' THEN 1 ELSE 0 END AS cart_pg,
            CASE WHEN wp.pageview_url = '/shipping' THEN 1 ELSE 0 END AS shipping_pg,
            CASE WHEN wp.pageview_url = '/billing' THEN 1 ELSE 0 END AS billing_pg,
            CASE WHEN wp.pageview_url = '/thank-you-for-your-order' THEN 1 ELSE 0 END AS thank_you_pg
        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'
            AND wp.created_at BETWEEN '2012-08-05' AND '2012-09-05'
        ORDER BY
            1,
            2
    ) AS pageview_level
GROUP BY
    1;
```

Click-through flags for each of **4,493** sessions

| website_session_id | made_to_products | made_to_fuzzy | made_to_cart | made_to_shipping | made_to_billing | made_to_thank_you |
|---|---|---|---|---|---|---|
| 18243 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18244 | 1 | 1 | 1 | 1 | 1 | 0 |
| 18245 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18246 | 1 | 0 | 0 | 0 | 0 | 0 |

```sql
-- Step 2: Finding the session_counts to a particular page
SELECT
    COUNT(DISTINCT website_session_id) AS total_sessions,
    COUNT( CASE WHEN made_to_products = 1 THEN website_session_id ELSE NULL END ) AS lander1_to_products,
    COUNT( CASE WHEN made_to_fuzzy = 1 THEN website_session_id ELSE NULL END ) AS products_to_fuzzy,
    COUNT( CASE WHEN made_to_cart = 1 THEN website_session_id ELSE NULL END ) AS fuzzy_to_cart,
    COUNT( CASE WHEN made_to_shipping = 1 THEN website_session_id ELSE NULL END ) AS cart_to_shipping,
    COUNT( CASE WHEN made_to_billing = 1 THEN website_session_id ELSE NULL END ) AS shipping_to_billing,
    COUNT( CASE WHEN made_to_thank_you = 1 THEN website_session_id ELSE NULL END ) AS billing_to_thank_you
FROM
    session_level_flags;
```

| total_sessions | lander1_to_products | products_to_fuzzy | fuzzy_to_cart | cart_to_shipping | shipping_to_billing | billing_to_thank_you |
|---|---|---|---|---|---|---|
| 4493 | 2115 | 1567 | 683 | 455 | 361 | 158 |

Conversion **tends to go down** as a user **moves forward in the conversion funnel**!

```sql
-- Step 3: Finding CTRs between each step in the conversion funnel
SELECT
    COUNT(DISTINCT website_session_id) AS total_sessions,
    ROUND( COUNT(CASE WHEN made_to_products = 1 THEN website_session_id ELSE NULL END) /
                COUNT(DISTINCT website_session_id) * 100, 2 ) AS lander1_ctr,
    ROUND( COUNT(CASE WHEN made_to_fuzzy = 1 THEN website_session_id ELSE NULL END) /
                COUNT(CASE WHEN made_to_products = 1 THEN website_session_id ELSE NULL END) * 100, 2 ) AS lander1_to_products_ctr,
    ROUND( COUNT(CASE WHEN made_to_cart= 1 THEN website_session_id ELSE NULL END) /
                COUNT(CASE WHEN made_to_fuzzy = 1 THEN website_session_id ELSE NULL END) * 100, 2 ) AS products_to_fuzzy_ctr,
    ROUND( COUNT(CASE WHEN made_to_shipping = 1 THEN website_session_id ELSE NULL END) /
                COUNT(CASE WHEN made_to_cart= 1 THEN website_session_id ELSE NULL END) * 100, 2 ) AS fuzzy_to_cart_ctr,
    ROUND( COUNT(CASE WHEN made_to_billing = 1 THEN website_session_id ELSE NULL END) /
                COUNT(CASE WHEN made_to_shipping = 1 THEN website_session_id ELSE NULL END) * 100, 2 ) AS cart_to_shipping_ctr,
    ROUND( COUNT(CASE WHEN made_to_thank_you = 1 THEN website_session_id ELSE NULL END) /
                COUNT(CASE WHEN made_to_billing = 1 THEN website_session_id ELSE NULL END) * 100, 2 ) AS shipping_to_billing_ctr
FROM
    session_level_flags;
```

| total_sessions | lander1_ctr | lander1_to_products_ctr | products_to_fuzzy_ctr | fuzzy_to_cart_ctr | cart_to_shipping_ctr | shipping_to_billing_ctr |
|---|---|---|---|---|---|---|
| 4493 | 47.07 | 74.09 | 43.59 | 66.62 | 79.34 | 43.77 |

We should focus on improving the performance of the new **/lander-1**, **/the-original-mr-fuzzy**, and the **/billing** pages which have the **lowest click-through rates.**

**NEW MESSAGE**
*November 10, 2012*

From: **Morgan Rockwell** *(Website Manager)*
Subject: **Conversion Funnel Test Results**

Hello!

We **tested an updated billing page** based on your funnel analysis. Can you take a look and see whether **/billing-2** is doing any better than the original **/billing** page?

We're wondering **what % of sessions on those pages end up placing an order**. FYI – we ran this test for **all traffic, not just for our search visitors**.

Thanks!
-Morgan

Reply Forward

```sql
/* Conversion Funnel Test Results: billing-2 vs billing till '2012-11-10' */
SELECT
    wp.pageview_url,
    COUNT(DISTINCT wp.website_session_id) AS sessions,
    COUNT(DISTINCT o.order_id) AS orders,
    COUNT(DISTINCT o.order_id) /
        COUNT(DISTINCT wp.website_session_id) to_order_pct
FROM
    website_pageviews wp
    LEFT JOIN orders o ON wp.website_session_id = o.website_session_id
WHERE
    wp.created_at < '2012-11-10'
    AND wp.pageview_url IN ('/billing', '/billing-2')
    AND wp.website_pageview_id > (
        SELECT MIN(website_pageview_id) -- first time 'b2' was seen
        FROM website_pageviews
        WHERE pageview_url = '/billing-2'
    ) -- output = 53550
GROUP BY
    1
;
```

| pageview_url | sessions | orders | to_order_pct |
|---|---|---|---|
| /billing | 657 | 300 | 0.4566 |
| /billing-2 | 653 | 409 | 0.6263 |

The new version of billing page '**/billing-2**' is definitely a **success** as the **billing-to-order rate** is **improved by** more than **17%,** and thus, the engineering team should **rollout this new version for entire traffic**.

# Thank You

Md Tauhid Alam

LinkedIn: https://www.linkedin.com/in/mdtauhidalam9231/