



SQL Capstone: Attribution

Learn SQL from Scratch

Brit Eide

28 June 2018

Table of Contents

1. Get familiar with CoolTShirts

- a. How many campaigns and sources does CoolTShirts use?
- b. How are they related?
- c. What pages are on the CoolTShirts website?

2. What is the user journey?

- a. How many first touches is each campaign responsible for?
- b. How many last touches is each campaign responsible for?
- c. How many visitors make a purchase?
- d. How many last touches on the purchase page is each campaign responsible for?
- e. What is the typical user journey?

3. Optimize the campaign budget

CoolTShirts can re-invest in 5 campaigns. Which should they pick, and why?

1. Get familiar with CoolTShirts

1a How many campaigns and sources does CoolTShirts use?

Number of Campaigns	Number of Sources
8	6

- There are 8 different campaigns on the CoolTShirts website
- There are 6 different sources

```
-- 1a
/* Campaigns */
SELECT count(distinct utm_campaign) AS Campaign
FROM page_visits; -- 8 campaigns

/* Sources */
SELECT count(distinct utm_source) AS Source
FROM page_visits; -- 6 sources

-- 1b
/* Source by Campaign */
SELECT distinct(utm_campaign) AS Campaign,
               utm_source AS Source
FROM page_visits
GROUP BY 1;
```

1b How are the campaigns and sources related?

The **Campaign** is what brought visitors to the site, and **Source** is how they got there.

For example, the Newsletter brings quite a few people to the CoolTShirts website. The newsletter is sent out by email, so this is *how* the newsletter customers are coming to the site. The newsletter is always sent by email, so the *source* for 'Weekly Newsletter' will therefore always be Email. Each campaign has one source, as shown below:

Campaign	Source
CoolTShirts Search	Google
Getting to know CoolTShirts	NY Times
Interview with CoolTShirts Founder	Medium
Paid Search	Google
Retargeting Ad	Facebook
Retargeting Campaign	Email
Ten Crazy CoolTShirts Facts	Buzzfeed
Weekly Newsletter	Email

1c What pages are on the CoolTShirts website?

Page Name
1 - Landing Page
2 - Shopping Cart
3 - Checkout
4 - Purchase

```
/* Distinct page */  
SELECT DISTINCT page_name AS 'Page Name'  
FROM page_visits;
```

2. What is the user journey?

2a How many First Touches is each campaign responsible for?

Source	Campaign	First Touch
Medium	Interview with CoolTshirts Founder	622
NY Times	Getting to Know CoolTShirts	612
Buzzfeed	Ten Crazy CoolTShirts Facts	576
Google	CoolTShirts Search	169

2b How many Last Touches is each campaign responsible for?

Source	Campaign	Last Touch
Email	Weekly newsletter	447
Facebook	Retargeting ad	443
Email	Retargeting campaign	245
NY Times	Getting to know CoolTShirts	232
Buzzfeed	Ten Crazy Cool Tshirt Facts	190
Medium	Interview with CoolTShirts founder	184
Google	Paid search	178
Google	CoolTShirts search	60

Queries for First and Last Touches

```
/* First Touches */
WITH first_touch_at AS (
    SELECT user_id,min(timestamp) AS
    'first_touch_at'
    FROM page_visits
    GROUP BY user_id),
ft_attr AS (
    SELECT ft.user_id,
    ft.first_touch_at,
    pv.utm_source,
    pv.utm_campaign
    FROM first_touch_at ft
    JOIN page_visits pv
        ON ft.user_id = pv.user_id
        AND ft.first_touch_at =
pv.timestamp)
SELECT ft_attr.utm_source AS Source,
    ft_attr.utm_campaign AS Campaign,
    COUNT(*) AS 'First Touch'
FROM ft_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

```
/* Last Touches */
WITH last_touch_at AS (
    SELECT user_id,max(timestamp) AS
    'last_touch_at'
    FROM page_visits
    GROUP BY user_id),
lt_attr AS (
    SELECT lt.user_id,
    lt.last_touch_at,
    pv.utm_source,
    pv.utm_campaign
    FROM last_touch_at lt
    JOIN page_visits pv
        ON lt.user_id = pv.user_id
        AND lt.last_touch_at =
pv.timestamp)
SELECT lt_attr.utm_source as Source,
    lt_attr.utm_campaign AS Campaign,
    COUNT(*) AS 'Last Touch'
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

2c How many visitors make a purchase?

Visitors	Percentage
361	18.24

Of the 1979 visitors to the CoolTShirts website, 361 - or just over 18% - made a purchase.

```
/* Visitors making a purchase */
SELECT COUNT(DISTINCT user_id) AS
Purchases,
      ROUND(100.0 *
            COUNT(DISTINCT user_id) /
            (SELECT COUNT(DISTINCT user_id)
             FROM page_visits)
            ,2) AS Percentage
FROM page_visits
WHERE page_name LIKE '%purchase';
```

2d How many last touches on the purchase page is each campaign responsible for?

Source	Campaign	Purchases
Email	Weekly newsletter	115
Facebook	Retargeting ad	113
Email	Retargeting campaign	54
Google	Paid search	52
Buzzfeed	Getting to know CoolTShirts	9
NY Times	Ten Crazy Cool TShirt Facts	9
Medium	Interview with CoolTShirts founder	7
Google	CoolTShirts search	2

The most successful campaign has been the **Weekly Newsletter**, with 115 purchases.

```
/* Last touch on the purchase page */
WITH last_touch_at AS (
  SELECT user_id,
    MAX(timestamp) AS 'last_touch_at'
  FROM page_visits
  WHERE page_name LIKE '%purchase'
  GROUP BY user_id),
lt_attr AS (
  SELECT lt.user_id,
    lt.last_touch_at,
    pv.utm_source,
    pv.utm_campaign
  FROM last_touch_at lt
  JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp)
SELECT lt_attr.utm_source AS Source,
  lt_attr.utm_campaign AS Campaign,
  COUNT(*) AS 'Last Touch'
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

2e What is the typical customer journey?

Most visitors with who come to the site on the landing page ('1 - Landing page') go through to the shopping cart, but the majority then seem to leave the site; only around 2% of customers complete their purchases.

The targeted campaigns - the ones where the customers are brought straight to the checkout page - do a lot better: nearly a quarter of all visitors complete their purchase.

```
/* Customer journey */  
SELECT utm_campaign AS Campaign,  
       page_name AS 'Page name',  
       COUNT(timestamp) AS 'Visitors'  
FROM page_visits  
GROUP BY 1,2 ORDER BY 1,2 ASC;  
-- percentages shown in next query
```

The next two slides show the customer journey broken down by campaign and page name:

Campaign	Page name	Visitors
CoolTShirts search	1 - Landing page	171
CoolTShirts search	2 - Shopping cart	133
CoolTShirts search	3 - Checkout	7
CoolTShirts search	4 - Purchase	2
Getting to know CoolTShirts	1 - Landing page	617
Getting to know CoolTShirts	2 - Shopping cart	682
Getting to know CoolTShirts	3 - Checkout	41
Getting to know CoolTShirts	4 - Purchase	9
Interview with CoolTShirts Founder	1 - Landing page	625
Interview with CoolTShirts Founder	2 - Shopping cart	515
Interview with CoolTShirts Founder	3 - Checkout	31
Interview with CoolTShirts Founder	4 - Purchase	7

Campaign	Page name	Visitors
Paid Search	3 - Checkout	179
Paid Search	4 - Purchase	52
Retargetting ad	3 - Checkout	445
Retargetting ad	4 - Purchase	113
Retargetting Campaign	3 - Checkout	246
Retargetting Campaign	4 - Purchase	54
Ten Crazy CoolTShirts Facts	3 - Checkout	32
Ten Crazy CoolTShirts Facts	4 - Purchase	9
Ten Crazy CoolTShirts Facts	1 - Landing page	587
Ten Crazy CoolTShirts Facts	2 - Shopping cart	570
Ten Crazy CoolTShirts Facts	3 - Checkout	32
Ten Crazy CoolTShirts Facts	4 - Purchase	9

3. Optimize the campaign budget

CoolTShirts can re-invest in 5 campaigns.

Which should they pick, and why?

The four most successful campaigns are clear, but in joint 5th place are 'Getting to know CoolTShirts' and 'Ten Crazy CoolTShirts Facts'. Which of these campaigns should CoolTShirts reinvest in?

Looking at the number of purchases in more detail, we see that the percentage of 'Ten Crazy CoolTShirts Facts' conversions is slightly higher than 'Getting to know CoolTShirts'.

It's pretty marginal, but this is the reason I am picking 'Ten Crazy CoolTShirts Facts' as the 5th campaign to reinvest in.

```
/* Purchases as % of total timestamps */
WITH count_campaigns AS (
    SELECT DISTINCT utm_campaign
    FROM page_visits),
percentage AS (
    SELECT DISTINCT utm_campaign,
    COUNT(timestamp) AS timestamp
    FROM page_visits
    WHERE page_name LIKE '%purchase'
    GROUP BY 1)
SELECT co.utm_campaign AS Campaign,
COUNT(pv.timestamp) AS 'Overall',
p.timestamp AS 'Purchases',
ROUND(100.0*
    p.timestamp/
    (COUNT(pv.timestamp))
    ,2) AS 'Percentage'
FROM page_visits pv
JOIN count_campaigns co
ON pv.utm_campaign=co.utm_campaign
JOIN percentage p
ON pv.utm_campaign=p.utm_campaign
GROUP BY 1
ORDER BY 4 DESC
LIMIT 5;
```

The 5 campaigns CoolTShirts should re-invest in by percentage purchased

Campaign	Overall	Purchases	Percentage
Paid Search	231	52	22.51%
Weekly Newsletter	565	115	20.35%
Retargetting Ad	558	113	20.25%
Retargetting Campaign	300	54	18%
Ten Crazy CoolTShirts Facts	1198	9	0.75%