



Capstone: Attribution

Learn SQL from Scratch

Jonathan Toomey

July 2018

Table of Contents

1. Get familiar with CoolTShirts
2. What is the user journey?
3. Optimize the campaign budget

1. Get familiar with CoolTShirts

1.1 How many campaigns and sources do CoolTShirts use?

CoolTShirts have 8 distinct campaigns and 6 distinct sources, as listed by the query shown.

utm_campaign is the method by which a user has been redirected to the CoolTShirts page: e.g. an internet search, an advert or a weekly newsletter.

utm_source is the medium in which the campaign was active when the user was redirected by it: e.g. a NY Times article, facebook, google, email or buzzfeed

```
SELECT COUNT(DISTINCT utm_campaign) AS 'Number of campaigns',  
COUNT(DISTINCT utm_source) AS 'Number of sources'  
FROM page_visits;
```

Number of Campaigns	Number of sources
8	6

1.2 How are these campaigns and sources related?

The campaigns are active in specific sources, as listed by the query shown.

As an example, the ‘getting to know CoolTShirts’ article (the campaign) was published in the NY Times (the source)

The data shows that the searches made were all conducted on Google; one such source being a ‘paid search’ and the other a ‘cool tshirts search’. There were 2 campaigns delivered by email: one retargeting, and one through the distribution of a weekly newsletter; presumably previous customers and/or subscribers.

Additionally, a ‘ten crazy CoolTShirts facts’ piece was posted on buzzfeed, and an interview with the CoolTShirts founder was published on ‘medium’. An advertising campaign took place on facebook.

The results from this query are consistent with the numbers presented previously in section 1.1.

```
SELECT DISTINCT utm_campaign AS 'Campaign', utm_source AS  
'Source'  
FROM page_visits  
GROUP BY 1;
```

Campaign	Source
cool-tshirts-search	google
getting-to-know-cool-tshirts	nytimes
interview-with-cool-tshirts-founder	medium
paid-search	google
retargeting-ad	facebook
retargeting-campaign	email
ten-crazy-cool-tshirts-facts	buzzfeed
weekly-newsletter	email

1.3 What Pages are on their website?

There are only 4 pages tracked in the database from the CoolTShirts website:

The landing page,
the shopping cart,
the checkout
and the purchase page

```
SELECT DISTINCT page_name AS 'Pages'  
FROM page_visits;
```

Pages
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

2. What is the user journey?

2.1 How many first touches is each campaign responsible for?

The query shown reports that the article on 'medium', the article in the NY Times and the feature on buzzfeed were overwhelmingly responsible for the number of 'first touches' on the CoolTShirts website. Google accounted for only 9% of the 1979 first touches.

```
/* Number of first touches for each campaign
*/
WITH first_touch AS (
  SELECT user_id,
    MIN(timestamp) as first_touch_at
  FROM page_visits
  GROUP BY user_id),

ft_attributed AS (
  SELECT ft.user_id,
    ft.first_touch_at,
    pv.utm_source,
    pv.utm_campaign
  FROM first_touch ft
  JOIN page_visits pv
  ON ft.user_id = pv.user_id
  AND ft.first_touch_at = pv.timestamp
)

SELECT fta.utm_source AS 'Source',
  fta.utm_campaign AS 'Campaign',
  COUNT(*) AS 'Count'
FROM ft_attributed fta
GROUP BY 2
ORDER BY 3 DESC;
```

Source	Campaign	Count of 'last touches'	Percentage of total
medium	interview-with-cool-tshirts-founder	622	31%
nytimes	getting-to-know-cool-tshirts	612	31%
buzzfeed	ten-crazy-cool-tshirts-facts	576	29%
google	cool-tshirts-search	169	9%

2.2 How many last touches is each campaign responsible for?

The query shown reports that email accounted for 35% of all last touches, with the facebook advert accounting for 22%. In contrast to the first touches, the articles in the NY Times and on buzzfeed and medium only account for 31% between them. Again, google is lowest, having only accounted for 12% of the 1979 overall last touches on the CoolTShirts website

```
/* Number of last touches for each campaign
*/
WITH last_touch AS (
  SELECT user_id,
    MAX(timestamp) as last_touch_at
  FROM page_visits
  GROUP BY user_id),

It_attributed AS (
  SELECT It.user_id,
    It.last_touch_at,
    pv.utm_source,
    pv.utm_campaign
  FROM last_touch It
  JOIN page_visits pv
    ON It.user_id = pv.user_id
    AND It.last_touch_at = pv.timestamp
)

SELECT Ita.utm_source AS 'Source',
  Ita.utm_campaign AS 'Campaign',
  COUNT(*) AS 'Count'
FROM It_attributed Ita
GROUP BY 2
ORDER BY 3 DESC;
```

Source	Campaign	Count of 'last touches'	Percentage of total
email	weekly-newsletter	447	23%
facebook	retargetting-ad	443	22%
email	retargetting-campaign	245	12%
nytimes	getting-to-know-cool-tshirts	232	12%
buzzfeed	ten-crazy-cool-tshirts-facts	190	10%
medium	interview-with-cool-tshirts-founder	184	9%
google	paid-search	178	9%
google	cool-tshirts-search	60	3%

2.3 How many visitors make a purchase?

The query shown reports that out of 1979 visitors to the CoolTShirts website, only 361 made a purchase.

This is a 'conversion rate' of only 18%

```
/*  
Number of visitors making a purchase  
*/  
SELECT COUNT(DISTINCT(user_id)) as 'Number of visitors making a purchase'  
FROM page_visits  
WHERE page_name = '4 - purchase';
```

Number of visitors making a purchase

361

2.4 How many last touches *on the purchase page* is each campaign responsible for?

The query shown reports that email accounted for 47% of all last touches on the purchase page, with the facebook advert accounting for 31%. This agrees with the attribution of last touches to each source/campaign and shows a high 'conversion rate' for these sources/campaigns.

Although google searches only accounted for 12% of last touches, it accounts for 15% of the purchase page last touches, suggesting that the 'conversion rate' is very good. The majority of successful google conversions are attributed to the paid search.

In stark contrast to the first touch data, the articles in the NY Times and on medium and buzzfeed only account for a total of 6% of the last touches on the purchase page.

```
/*
Number of last touches ON THE PURCHASE PAGE
for each campaign
*/
WITH last_touch AS (
  SELECT user_id,
    MAX(timestamp) as last_touch_at
  FROM page_visits
  WHERE page_name = '4 - purchase'
  GROUP BY user_id),

It_attributed AS (
  SELECT It.user_id,
    It.last_touch_at,
    pv.utm_source,
    pv.utm_campaign,
    pv.page_name
  FROM last_touch It
  JOIN page_visits pv
  ON It.user_id = pv.user_id
  AND It.last_touch_at = pv.timestamp
)
SELECT Ita.utm_source AS 'Source',
  Ita.utm_campaign AS 'Campaign',
  COUNT(*) AS 'Count'
FROM It_attributed Ita
GROUP BY 2
ORDER BY 3 DESC;
```

Source	Campaign	Count of 'last touches'	Percentage of total
email	weekly-newsletter	115	32%
facebook	retargetting-ad	113	31%
email	retargetting-campaign	54	15%
google	paid-search	52	14%
nytimes	getting-to-know-cool-tshirts	9	2%
buzzfeed	ten-crazy-cool-tshirts-facts	9	2%
medium	interview-with-cool-tshirts-founder	7	2%
google	cool-tshirts-search	2	1%

2.5 What is the typical user journey?

Based on the above data, the typical user journey (using the highest percentage values) is as follows:

Most of the first touches are generated through reading the articles on buzzfeed, medium or the NY Times. Of these 1813 first touches, only 1.4% (25) go on to make a purchase.

The majority of these users leave the website without making a purchase.

These users are then targeted with an email or facebook advert, which prompts them to revisit the CoolTShirts website.

The email and facebook campaigns account for 57% of the last touches. Of these 1135 last touches, 25% (282) are converted into a purchase.

Source	Campaign	Count of 'first touches'	Percentage of total (1979)	Count of 'last touches'	Percentage of total (1979)	Count of 'last touches' on purchase page	Percentage of total (361)	Conversion as Percentage of first touches	Conversion as Percentage of last touches
email	weekly-newsletter			447	23%	115	32%		26%
facebook	retargeting-ad			443	22%	113	31%		26%
email	retargeting-campaign			245	12%	54	15%		22%
google	paid-search			178	9%	52	14%		29%
nytimes	getting-to-know-cool-tshirts	612	31%	232	12%	9	2%	1.5%	4%
buzzfeed	ten-crazy-cool-tshirts-facts	579	29%	190	10%	9	2%	1.6%	5%
medium	interview-with-cool-tshirts-founder	622	31%	184	9%	7	2%	1.1%	4%
google	cool-tshirts-search	169	9%	60	3%	2	1%	1.2%	3%

3. Optimize the campaign budget

3.1 Which campaigns and why?

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

Of the users to visit CoolTShirts, the vast majority come from the NY Times, buzzfeed and medium. However, very few of these users made a purchase until they receive a follow-up targeted email or facebook advert.

If CoolTShirts want to encourage more first touches, then continuing to invest in adverts/articles on NY Times, buzzfeed and medium.

However, in order to convert these first touches into purchases, CoolTShirts need to continue investing in their email, facebook and paid search google campaigns; accounting as they do for 92.5% of the last touches on the purchase page.

Therefore my recommendation would be to continue the campaigns in the NY Times and buzzfeed in order to generate new users (these account for 60% of first touches), as they represent the best conversion from first touch to purchase.

CoolTShirts should also continue to retarget users with the weekly newsletter and advert by email, and the targeted advertising on facebook campaigns as these account for 78% of the last touches on the purchase page.

However, I would strongly recommend CoolTShirts to consider attempting to raise additional funds in order to also continue the paid search campaign on google, which accounts for 14% of their sales.

Source	Campaign
email	weekly-newsletter
facebook	retargetting-ad
email	retargetting-campaign
nytimes	getting-to-know-cool-tshirts
buzzfeed	ten-crazy-cool-tshirts-facts