



Capstone Project: First- and Last-Touch Attribution with CoolTShirts.com

Learn SQL from Scratch

Maria Sarafi

Date: July 17th, 2018

Table of Contents

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

I. Get familiar with CoolTShirts

1.1 How many campaigns and sources does CoolTShirts use and how are they related?

CoolTShirts is an innovative apparel shop that sells cool and creatively designed T-shirts. Recently, CTS started a few marketing campaigns to increase website visits and purchases. CTS is running 8 campaigns through 6 traffic sources (sometimes called channels or touchpoints) online. Let's examine what is the difference between the terms `utm_campaign` and `utm_source` and how these two terms are related:

- The term `utm_campaign` identifies the specific ad or email blast (i.e., Summer sale newsletter, 'Back to school promotional code' ad, Facebook ads) put forward by a company to drive engagement, conversions, traffic, or revenue.
- The term `utm_source` identifies which site sent the traffic (i.e., Google, Facebook etc).

In the table on the right, we can see the marketing campaigns and the sources and how they are related.

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

I have used three queries:

- one for the number of distinct campaigns,
- one for the number of distinct sources,
- one to find how they are related.

```
SELECT COUNT(DISTINCT utm_campaign)
FROM page_visits;
```

```
SELECT
COUNT(DISTINCT utm_source)
FROM page_visits;
```

```
SELECT DISTINCT utm_campaign, utm_source
FROM page_visits;
```

1.2 What pages are on the CoolTShirts website?

The CoolTShirts website consists of four pages: landing page, shopping cart, checkout and purchase. They are shown in the table below. The query used is in the table on the right.

page_name
landing_page
shopping_cart
checkout
purchase

```
SELECT DISTINCT page_name  
FROM page_visits;
```

2. What is the user journey?

2.1 How many first touches is each campaign responsible for?

The number of first touches (the first time the users are exposed to CoolTShirts.com) for each campaign is shown in the table below. The campaigns that are not in the table are responsible for visits but not for first touches.

ft_attr.utm_source	ft_attr.utm_campaign	COUNT(*)
medium	interview-with-cool-tshirts-founder	622
nytimes	getting-to-know-cool-tshirts	612
buzzfeed	ten-crazy-cool-tshirts-facts	576
google	cool-tshirts-search	169

```
WITH first_touch 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 ft  
  JOIN page_visits pv  
    ON ft.user_id = pv.user_id  
   AND ft.first_touch_at = pv.timestamp)  
SELECT ft_attr.utm_source,  
       ft_attr.utm_campaign,  
       COUNT(*)  
FROM ft_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```


2.2 How many last touches is each campaign responsible for?

The number of last touches (the exposure to CoolTShirts.com that led to a purchase) for each campaign is shown in the table on the right.

```
WITH last_touch 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 lt  
  JOIN page_visits pv  
    ON lt.user_id = pv.user_id  
   AND lt.last_touch_at = pv.timestamp)  
SELECT lt_attr.utm_source,  
       lt_attr.utm_campaign,  
       COUNT(*)  
FROM lt_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

lt_attr.utm_source	lt_attr.utm_campaign	COUNT(*)
email	weekly-newsletter	447
facebook	retargetting-ad	443
email	retargetting-campaign	245
nytimes	getting-to-know-cool-tshirts	232
buzzfeed	ten-crazy-cool-tshirts-facts	190
medium	interview-with-cool-tshirts-founder	184
google	paid-search	178
google	cool-tshirts-search	60

2.3 How many visitors make a purchase?

The number of distinct users who visited the page named '4 - purchase' is 361. The query used is in the table on the right.

```
SELECT COUNT(DISTINCT(user_id))  
FROM page_visits  
WHERE page_name = '4 - purchase';
```

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

We can calculate the number of last touches on the purchase page by adding a WHERE clause to the last-touch query.

```
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),  
lt_attr AS (  
  SELECT lt.user_id,  
         lt.last_touch_at,  
         pv.utm_source,  
         pv.utm_campaign  
  FROM last_touch lt  
  JOIN page_visits pv  
    ON lt.user_id = pv.user_id  
   AND lt.last_touch_at = pv.timestamp)  
SELECT lt_attr.utm_source,  
       lt_attr.utm_campaign,  
       COUNT(*)  
FROM lt_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

lt_attr.utm_source	lt_attr.utm_campaign	COUNT(*)
email	weekly-newsletter	115
facebook	retargetting-ad	113
email	retargetting-campaign	54
google	paid-search	52
buzzfeed	ten-crazy-cool-tshirts-facts	9
nytimes	getting-to-know-cool-tshirts	9
medium	interview-with-cool-tshirts-founder	7
google	cool-tshirts-search	2

2.5 What is the typical user journey?

In order to map the user journey and understand its flow, we can use the query in the right table. The results are shown in the table on the right (below the query).

The users typically start browsing for T-shirts on the landing page, then they add the products they like to their shopping cart, proceed to the checkout page where they enter any required personal information and paying details and finally they purchase the desired products.

We realise that between the third and fourth step there is a huge cart abandonment problem of 74.77%. Understanding what leads different customers to bail out on a purchase is a crucial step to fixing any gaps in the user journey and converting more of the website visitors to happy purchasers.

```
SELECT page_name, COUNT(DISTINCT(user_id))  
FROM page_visits  
GROUP BY page_name;
```

page_name	COUNT(DISTINCT(user_id))
landing_page	1979
shopping_cart	1881
checkout	1431
purchase	361

3. Optimize the campaign budget

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

CoolTShirts can re-invest in 5 campaigns and we should help them pick the most effective ones.

We are going to examine the conversion (the percentage of visitors of the website that successfully convert into leads) for each campaign. First, we will calculate the conversion from checkout to purchase. The results are shown in the next table.

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

Source	Campaign	Last touches (on any page)	Last touches (on purchase page)	Conversion rate
email	weekly-newsletter	447	115	0.257
facebook	retargetting-ad	443	113	0.255
email	retargetting-campaign	245	54	0.220
google	paid-search	232	52	0.224
buzzfeed	ten-crazy-cool-tshirts-facts	190	9	0.047
nytimes	getting-to-know-cool-tshirts	184	9	0.049
medium	interview-with-cool-tshirts-founder	178	7	0.039
google	cool-tshirts-search	60	2	0.033

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

Then, we will calculate the conversion from first touch to purchase. The results are shown in the table below.

Source	Campaign	First touches (on any page)	Last touches (on purchase page)	Conversion rate
buzzfeed	ten-crazy-cool-tshirts-facts	576	9	0.016
nytimes	getting-to-know-cool-tshirts	612	9	0.015
medium	interview-with-cool-tshirts-founder	622	7	0.011
google	cool-tshirts-search	169	2	0.012

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

The most important goal of an online store is to generate revenue. It is obvious that the most successful campaigns (the ones with the higher conversion rate) that generate the higher sales are: the weekly newsletter, the Facebook retargeting ad, Google's paid search and the Email retargeting campaign.

On the other hand, as we see in the second table, the campaigns with the lower conversion rate bring many new visitors to the website and therefore increase the website traffic and build brand awareness around CoolTShirts.

For these reasons, I would suggest that CoolTshirts should re-invest in the four campaigns that resulted in the higher number of transactions and in one of the other campaigns that don't have a high conversion rate but can bring a new audience to the CoolTshirts website and increase the awareness around the brand and its products.