



# Attribution queries

Learn SQL from Scratch

Hannah Jenkins

July 17 2018

# Example 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 Campaigns and sources

First I looked at how many campaigns and sources CoolTShirts uses and established how are they related.

To do this I found out how many distinct page visits there were for both campaigns (utm\_campaign) – 8, and the source of the campaign (utm\_source) – 6. This provided a picture of what campaigns were being run and what source each campaign was associated with.

There are more campaigns than sources and I established that both email and google were used twice as sources for different campaigns.

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

```
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 Pages

I then took a look at the different pages of the website to establish the options for assessing the user journey.

I discovered there were four pages taking the user from the landing page, through to shopping car, then checkout and finally the purchase page.

Query Results
page_name
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

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

## **2. The user journey**

## 2.1 First touch

I then looked at how many first touches each campaign was responsible for.

To do this I queried the campaigns responsible for the user making the landing page, based on the minimum timestamp.

I then grouped them by campaign and counted the number of first touches for each.

I established that the interview with the CoolTshirts founder via the source medium was the leader, indicating this was a successful campaign at getting users to click through to the landing page.

Query Results

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 Last touch

I then queried for the last touch made by users to find out which campaigns were responsible for getting the user back to the website.

I established that the weekly newsletter was responsible for the most returns, however, by a small margin over the retargetting Facebook campaign ad.

Query Results		
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

```
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;
```



## 2.3 Purchases

I found out how many users made a purchase by querying for distinct users who made it to the purchase page. The number is 361.

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

### Query Results

COUNT (DISTINCT user_id)
--------------------------

361
-----

## 2.3 Purchases by campaign

I refined the last touch attribution query to find out how many last touches on the purchase page each campaign was responsible for.

It turned out there were two top campaigns, the weekly-newsletter campaign and the retargeting-ad via Facebook.

.

Query Results		
lt_attr.utm_source	lt_attr.utm_campaign	visits
email	weekly-newsletter	115
facebook	retargeting-ad	113
email	retargeting-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

```
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(*) AS visits  
FROM lt_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

## 2.5 User journey

The previous queries have revealed some salient things about the user journey for the CoolTShirts customers.

- Initially there were two stand out campaigns that got the user to the CoolTShirt website: Interview with CoolTShirts Founder and Getting to know the CoolTShirts via Medium and New York Times respectively.
- The last touch query showed that the weekly newsletter email and the Facebook ad retargeting the users were largely responsible for getting users back to the website.
- When it came to actual purchases, the same two campaigns also rated highly.

### **3. Optimize the campaign budget**

## 3.1 CoolTShirts future campaigns

The analysis indicated that CoolTShirts re-targeting campaigns are a successful way to compel the user to make a purchase they may have considered earlier after having their initial interest peaked via the campaigns that utilised general broadsheet information (the interview with the CEO and the Getting to Know CoolTshirts campaigns).

This indicates the five campaigns that CoolTshirts should reinvest in are:

- Interview with the CEO (or follow up of something similar)
  - Getting to Know CoolTShirts (or follow up of something similar)
  - Weekly newsletter (this scored high on both last touch and purchase conversion)
  - Retargeting campaign – Facebook\*
  - Retargeting campaign – email\*
- 
- \*This assumes that they are retargeting consumers based on analysis of first touch data