



Capstone: Attribution

Learn SQL from Scratch

Christopher Attewell-Hughes

23/06/2018



Contents

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



1. Get familiar with CoolTShirts

- How many campaigns and sources does CoolTShirts use and how are they related?
Be sure to explain the difference between **utm_campaign** and **utm_source**.
- What pages are on their website?



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

With campaigns being the particular marketing message being used to drive customers to site, and the source being the medium through which this is achieved, we need to understand how many of each CoolTshirts uses. An initial query shows:

- 8 campaigns are currently being run
- Of these 8 campaigns, 6 sources are used to deliver these messages.
- Email is being used for two campaigns as is Google.

```
SELECT COUNT (DISTINCT utm_campaign) AS 'campaign_count'  
FROM page_visits;
```

```
SELECT COUNT (DISTINCT utm_source) AS 'source_count'  
FROM page_visits;
```

```
SELECT DISTINCT utm_campaign, utm_source  
FROM page_visits  
ORDER BY 2;
```

campaign_count
8

source_count
6

utm_campaign
ten-crazy-cool-tshirts-facts
weekly-newsletter
retargetting-campaign
retargetting-ad
paid-search
cool-tshirts-search
interview-with-cool-tshirts-founder
getting-to-know-cool-tshirts

utm_source
buzzfeed
email
email
facebook
google
google
medium
nytimes

1.2 What pages are on their website?

To understand the layout of the CoolTshirts website, we can SELECT the DISTINCT page names.

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

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

2. What is the user journey?

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



2.1 How many first touches is each campaign responsible for?

By using the MIN function to find a customer's first interaction with the site within the confines of a specific campaign.

We can then use COUNT to identify the volume of unique customers per campaign driven to site.

The query result to the right indicates that:

- Of the 8 campaigns and 6 sources, only 4 of each are responsible for driving traffic to site.
- The top three drivers are article-based as opposed to email campaigns or paid searches.
- Google searches does feature, but is heavily outweighed by drivers from sites which publish articles.

```
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 AS ft  
    JOIN page_visits AS 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_Drivers  
FROM ft_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

Source	Campaign	First_Touch_Drivers
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

2.2 How many last touches is each campaign responsible for?

By using the MAX function to find a customer's last interaction with the site within the confines of a specific campaign. We can then use COUNT again to identify the volume of unique customers per campaign driven to site.

The query result to the below indicates that:

- Emails and Facebook retargeting ads are the most effective at drawing customers back to site
- This is followed closely by retargeting emails and the 'getting-to-know-cool-tshirts' campaign

This suggests that directly engaging with a customer through email and social media is most effective at bringing customers back to site

```
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,  
       pv.page_name  
FROM last_touch AS lt  
JOIN page_visits AS 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_Drivers  
FROM lt_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

Source	Campaign	Last_Touch_Drivers
email	weekly-newsletter	447
facebook	retargeting-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

2.4 How many visitors make a purchase?

Using the COUNT and DISTINCT function in tandem, we can identify the number of unique users who visited the site and then the number who went on to view the purchase page.

With 1979 visits and 361 purchases, this equates to a 18% visitor conversion

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

```
SELECT COUNT (DISTINCT user_id) AS Visits
FROM page_visits;
```

Purchases

361

Visits

1979

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

In conjunction with MAX query, we can use WHERE to identify campaigns which drove customers towards the purchase page.

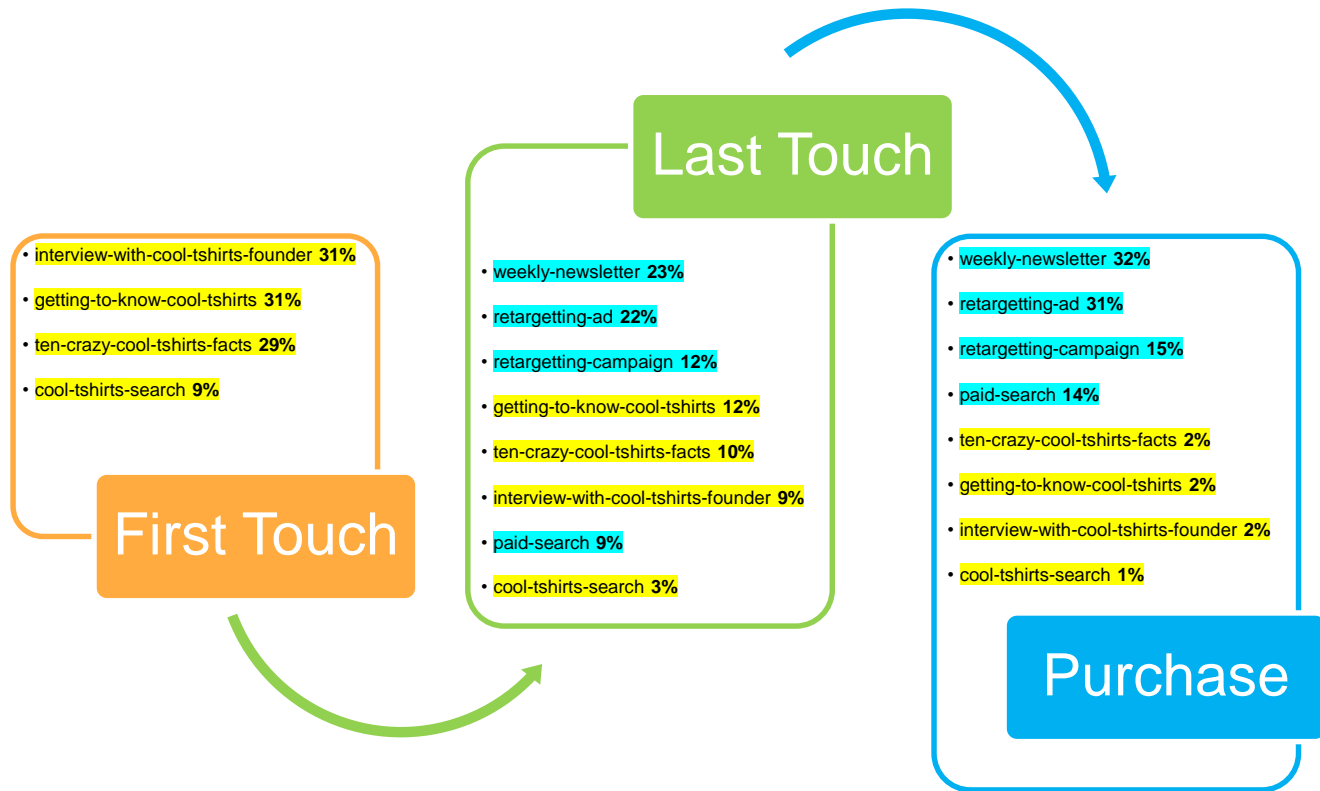
The query result to the below indicates that:

- Emailed newsletters and retargeting through email and Facebook are the most effective.
- Those which were most effective at bringing initial traffic to site are the least effective at driving customers to purchase.

```
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,  
           pv.page_name  
    FROM last_touch AS lt  
    JOIN page_visits AS 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 Purchase_Drivers  
FROM lt_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

Source	Campaign	Purchase_Drivers
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

2.5 What is the typical user journey?



- 91% of visitors are initially driven by article-based campaigns.
- Of these campaigns, only 34% accounts for a last touch
- Of all last touch instances which led to a purchase, article-based content accounts for 7%

- While email campaigns and retargeting ads don't drive the initial interest from customers, they do ensure a reasonable % return for a final purchase.

As such, emails and retargeting are vital to ensure the 18% conversion COOLTshirts currently enjoys.

3. Optimize the campaign budget

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



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

If COOLTshirts can only re-invest in 5 campaigns, it should choose:

- The 3 article-based campaigns (yellow) as these spur on the highest amount of initial traffic.

While they don't account for much in the way of purchases, they do still contribute.

- The emailed weekly-newsletter and the Facebook retargeting. This allows for COOLTshirts to approach those who have opted in to be contacted via email as well as those who have not (via Facebook).

Note: Ideally COOLTshirts would also elect to invest in more email retargeting, but it should choose the newsletter over this if it can only choose one as it has both a higher contribution to overall purchases as well as a higher last-touch to purchase conversion.

Source	Campaign	First_Touch_Drivers	Mix	Last_Touch_Drivers	Mix	Purchase_Drivers	Mix	Mix of Last_Touch
email	weekly-newsletter	0	0%	447	23%	115	32%	26%
facebook	retargeting-ad	0	0%	443	22%	113	31%	26%
email	retargeting-campaign	0	0%	245	12%	54	15%	22%
nytimes	getting-to-know-cool-tshirts	612	31%	232	12%	9	2%	4%
buzzfeed	ten-crazy-cool-tshirts-facts	576	29%	190	10%	9	2%	5%
medium	interview-with-cool-tshirts-founder	622	31%	184	9%	7	2%	4%
google	paid-search	0	0%	178	9%	52	14%	29%
google	cool-tshirts-search	169	9%	60	3%	2	1%	3%
	Total	1979	100%	1979	100%	361	100%	18%