



Capstone - Learn SQL from Scratch

First- and Last-Touch Attribution project
- *CoolTShirts*

Xichen Susie Guo
27 August 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

- How many campaigns and sources does CoolTShirts use and how are they related?
- What pages are on their website?

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

CoolTShirts have **8 campaigns** and the campaigns are run via **6 different sources** - as shown in table 1 and generated by query 1. The campaigns are identified by 'utm_campaign', as shown in table 2, which specifies each campaign individually. Each campaign's corresponding *source* is also shown in table 2 under the column 'utm_source' which identifies the source that led a user to CoolITShirts.

For example, both 'weekly-newsletter' and 'retargetting campaign' have sources marked as 'email' in the 'utm_source' column— which means traffic from these two campaigns were sent to CoolTShirts via emails.

Number of campaigns	Number of sources
8	6

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

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

QUERY 1

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

QUERY 2

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

TABLE 2

What pages are on their website?

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

QUERY 3

The pages on CoolTShirts' website include: 1. landing page; 2. shopping cart; 3. checkout; and 4. purchase. These are identified using query 3 which shows the output of 'page_name' (as a column) in table 3: 'page_name' specifies the title of the section of the page that was visited, hence identifying the pages of the website.

TABLE 3

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

2. What is your 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?

How many first touches is each campaign responsible for?

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

TABLE 4

As shown in table 4 above, in descending order, the number of first touches each campaign is responsible for. The 'interview-with-cool-tshirts-founder' campaign is responsible for the most first touches, with 622 visitors. It is followed by the 'getting-to-know-cool-tshirts' campaign, run in the New York Times at 612, then 'ten-crazy-cool-tshirts-facts' campaign via buzzfeed at 576; with 'cool-tshirts-search' campaign via google coming last generating 169 first touches to CoolTShirts' website.

These results were generated by query 4.

```
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_campaign,  
       ft_attr.utm_source,  
       COUNT(*)  
FROM ft_attr  
GROUP BY 1  
ORDER BY 3 DESC;
```

QUERY 4

How many last touches is each campaign responsible for?

The number of last touches driven by each campaign is listed below in table 5; in descending order; the results shown here are generated by the codes in query 5. The 'weekly-newsletter' campaign run via email is the most effective in terms of last touches to CoolTShirts at a count of 447 visitors and the 'retargetting-ad' campaign is close behind in 2nd at 443. The 'cool-tshirts-search' campaign is the least effective at 60.

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

TABLE 5

```
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 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_campaign,  
       lt_attr.utm_source,  
       COUNT(*)  
FROM lt_attr  
GROUP BY 1  
ORDER BY 3 DESC;
```

QUERY 5

How many visitors make a purchase?

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

QUERY 6

Visitors who purchased
361

TABLE 6

361 visitors made purchases on CoolTShirts' website. This result was the output of query 6, which counts the distinct users who visited the 'purchase' page on their website, hence capturing only those who purchased.

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

TABLE 7

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

Table 7 lists the 8 individual campaigns against the corresponding number of last touches they are responsible of visitors who were on the purchase page (3rd column), hence made purchases on CoolTShirts. The 'weekly-newsletter' campaign was the most effective in driving last touches that resulted in purchases, with 115 such last touches; 'cool-tshirts-search' was the least effective campaign in this with only 2.

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

Query 7 shows the codes that produced the output shown in table 7 on the previous slide.

```
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 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_campaign,  
       lt_attr.utm_source,  
       COUNT(*)  
FROM lt_attr  
GROUP BY 1  
ORDER BY 3 DESC;
```

Query 7

What is the typical user journey?

First touches table

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

Last touches purchase table

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

A typical user journey starts with watching the ‘interview with cool t-shirts founder’, which attracted 622 visitors (over 31% of the total amount of first touches) to CoolTShirts’ website. The other common way for a potential user to find out about CoolTShirts is through the campaign of ‘getting to know cool-tshirts’ run in the New York Times. The two campaigns combined generated a significant portion, over 62%, of first touches traffic to their website.

The second column in the ‘Last touches purchase table’ lists the sources of each campaign responsible for last touches with purchases - the info here suggest that some users provided information such as email addresses to CoolTShirts which allowed them to retarget these users and attract them back to the CoolTShirts website to make purchases. The ‘weekly newsletter’ campaign, run via email, accounts for 115 last touches that made purchases - the highest of all campaigns and responsible for nearly 32% of all purchases made on CoolTShirts.

This suggests that the most typical user journey could be a user who visited the website after watching the interview with CoolTShirts’ founder, registered their email, went away without purchasing on the site but returned to CoolTShirts after reading one of its weekly newsletter emails and then made a purchase as a result.

3. Optimize the campaign budget

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

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

First touches table

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

Last touches purchase table

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

To achieve commercial success CoolTShirts would need campaigns to deliver on two objectives 1). to attract visitors to their website, measured in both 'first touches' and 'last touches' 2). and perhaps more importantly, for 'last touches' to turn into purchases. Hence the criteria to determine which 5 campaigns to reinvest in would be a combination of campaigns that were the a). most effective in terms of last touches with purchases ('Last touches purchase table) and also first touch traffic (First touches table).

From the perspective of 'last-touches-turn-purchases', the two campaigns '**weekly-newsletter**' and '**retargeting-ad**' were the most successful and should be re-invested as they brought in 115 and 113 visitors respectively - which is over 63% of last touches with purchases made. 'Retargeting-campaign' comes 3rd with 54 visitors; however, it is a big drop from the 2nd place campaign in this aspect (at 113) which implies its limited value to CoolTShirts.

When looking at the 2 tables together, the 4 campaigns that generated 'first touches' have a small gap in their counts in both tables except for bottom of the 4 campaigns, i.e. 'cool-tshirts-search'; and hence it should be excluded. It could be argued that the campaigns that give CoolTShirts 'first touches' are more significant in the sense that a 'first-touch' is the start of the customer journey and without which 'last touch' or 'purchases' would not be possible. Therefore, '**Interview-with-cool-tshirts-founder**', '**getting-to-know-cool-tshirts**' and '**ten-crazy cool-tshirts-facts**' should be re-invested - together they were responsible for over 91% of first touches.