



Capstone Project

Option 3:

First- and Last-Touch Attribution with CoolTShirts.com

Learn SQL from Scratch

Burc SEN , October 8, 2018

Table of Contents

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

1. Get familiar with the company

1.1. 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`.

1.2. What pages are on their website?

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

Query Results	
Distinct No. of Campaigns	
8	
Distinct No. of Sources	
6	
Campaigns	Sources
ten-crazy-cool-tshirts-facts	buzzfeed
weekly-newsletter	email
retargetting-campaign	email
retargetting-ad	facebook
paid-search	google
cool-tshirts-search	google
interview-with-cool-tshirts-founder	medium
getting-to-know-cool-tshirts	nytimes

utm_campaign: Identifies the specific ad or email blast i.e. paid-search

utm_source: Identifies which site sent the traffic i.e. google

The data set shows that some campaigns used the same type of communication (traffic). So the number of campaigns is higher than the number of sources.

```
SELECT count(DISTINCT utm_campaign) AS "Distinct No. of Campaigns"
FROM page_visits;
```

```
SELECT count(DISTINCT utm_source) AS "Distinct No. of Sources"
FROM page_visits;
```

```
SELECT DISTINCT utm_campaign AS "Campaigns",
                utm_source AS "Sources"
FROM page_visits
ORDER BY sources ASC;
```

1.2. What pages are on their website?

Query Results
Distinct No. of Page Names
4
Page Name
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

Column „page_name“ shows the title of the section of the page that was visited. In total 4 distinct pages.

```
SELECT count(DISTINCT page_name) AS "Distinct No. of  
Page Names"  
FROM page_visits;
```

```
SELECT DISTINCT page_name AS "Page Name"  
FROM page_visits  
ORDER BY Page_Name ASC;
```

2. What is the user journey?

- 2.1. How many first touches is each campaign responsible for?
- 2.2. How many last touches is each campaign responsible for?
- 2.3. How many visitors make a purchase?
- 2.4. How many last touches on the purchase page is each campaign responsible for?
- 2.5. What is the typical user journey?

2.1. How many first touches is each campaign responsible for?

Query Results		
Source	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

SUM 1,979

First-touch attribution only considers the first source and campaign for each customer. This is a good way of knowing how visitors initially discover a website. 4 out of the 8 campaigns were responsible for first touches. Most first touches were received by the the campaign interview-with-cool-thshirts-founder.

```
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 AS "Source",  
       ft_attr.utm_campaign AS "Campaign",  
       COUNT(*) AS "Count"  
FROM ft_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

2.2. How many last touches is each campaign responsible for?

Query Results		
Source	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

SUM 1,979

Last-touch attribution only considers the last source and campaign for each customer. This is a good way of knowing how visitors are drawn back to a website, especially for making a final purchase. All campaigns had last touches. Most last touches were received by the campaign weekly-newsletter.

```
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 AS "Source",  
       lt_attr.utm_campaign AS "Campaign",  
       COUNT(*) AS "Count"  
FROM lt_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```


2.3. How many visitors make a purchase?

Query Results
Purchase
361

```
SELECT count(DISTINCT user_id) AS "Purchase"  
FROM page_visits  
WHERE page_name = "4 - purchase";
```

In total 361 purchases were made.

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

Query Results		
Source	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

SUM 361

The campaign which led to the most purchases was weekly-newsletter.

```
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 AS "Source",  
       lt_attr.utm_campaign AS "Campaign",  
       COUNT(*) AS "Count"  
FROM lt_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

2.5. What is the typical user journey?

The customers got attracted first by three campaigns (91%). The number of those campaigns dropped to 31% (drawing back). And those campaigns led to 6% of total purchases. Emails and Facebook were most successful drawing back customers (57%). Those sources and corresponding campaigns led to 78% of total purchases

Sources	Campaigns	Attracted first	Attracted last	Purchases completed
medium	interview-with-cool-tshirts-founder	31%	9%	2%
nytimes	getting-to-know-cool-tshirts	31%	12%	2%
buzzfeed	ten-crazy-cool-tshirts-facts	29%	10%	2%
google	cool-tshirts-search	9%	3%	1%
google	paid-search	n.a.	9%	14%
facebook	retargeting-ad	n.a.	22%	31%
email	retargeting-campaign	n.a.	12%	15%
email	weekly-newsletter	n.a.	23%	32%

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

The below campaigns and sources are proposed to be picked as they contributed to 62% of the first site visitors and led to 82% of purchases.

Sources	Campaigns	Attracted first	Attracted last	Purchases completed
medium	interview-with-cool-tshirts-founder	31%	9%	2%
nytimes	getting-to-know-cool-tshirts	31%	12%	2%
facebook	retargeting-ad	n.a.	22%	31%
email	retargeting-campaign	n.a.	12%	15%
email	weekly-newsletter	n.a.	23%	32%