



Learn SQL from Scratch

Capstone Project

Lisa A.

2.8.19

CoolTShirts

First- and Last-Touch Attribution Analysis

Contents

1. What is CoolTShirts?
2. What is the user journey?
3. How can we optimize the campaign budget?

Getting Familiar with CTS: Current Campaigns & Sources

How many current campaigns are there and what are the sources for them? We use UTM (urchin tracking module) parameters to track clicks on web pages and emails. (Each click is another row added to a table.) We can use the three queries below to see that there are 8 campaigns (email or targeted ads) and 6 sources (or, touchpoints that results in users going to the site), NYTimes, Google, email, BuzzFeed, Medium & Facebook.

```
3  SELECT COUNT(DISTINCT utm_campaign)
4  FROM page_visits;
5
6  SELECT COUNT(DISTINCT utm_source)
7  FROM page_visits;
8
9  SELECT DISTINCT utm_campaign, utm_source
10 FROM page_visits;
11
```

Query Results	
COUNT(DISTINCT utm_campaign)	
8	
COUNT(DISTINCT utm_source)	
6	
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

Getting Familiar with CTS, con't.

Currently, CoolTShirts has four different pages: landing, shopping_cart, checkout, purchase. We can use this query to find that out →

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

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

User Journey

To begin, we can use queries to find out how users initially arrived at CTS (via which campaign and source) and what their journey was on the site, ie: did it result in a purchase? The initial visit is called “first-touch” and is discovered by this query which tells us what campaign is attributed to the user visiting the site→

```
WITH first_touch AS (  
    SELECT user_id,  
           MIN(timestamp) as  
           first_touch_at  
    FROM page_visits  
    GROUP BY user_id)  
SELECT ft.user_id,  
       ft.first_touch_at,  
       pv.utm_source,  
       pv.utm_campaign,  
       COUNT(utm_campaign)  
FROM first_touch ft  
JOIN page_visits pv  
    ON ft.user_id = pv.user_id  
   AND ft.first_touch_at =  
      pv.timestamp  
GROUP BY pv.utm_campaign  
ORDER BY 5 DESC;
```

Query Results

user_id	first_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
99990	2018-01-13 23:30:09	medium	interview-with-cool-tshirts-founder	622
99933	2018-01-25 00:04:39	nytimes	getting-to-know-cool-tshirts	612
99765	2018-01-04 05:59:46	buzzfeed	ten-crazy-cool-tshirts-facts	576
99684	2018-01-13 13:20:49	google	cool-tshirts-search	169

First-Touch Attribution

We can see exactly how users arrived at the website and which campaign and source brought them here. The interview on Medium brought 622 users to the site, then the NYTimes article brought 612 users and so forth. But, what happened from there?

Query Results				
user_id	first_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
99990	2018-01-13 23:30:09	medium	interview-with-cool-tshirts-founder	622
99933	2018-01-25 00:04:39	nytimes	getting-to-know-cool-tshirts	612
99765	2018-01-04 05:59:46	buzzfeed	ten-crazy-cool-tshirts-facts	576
99684	2018-01-13 13:20:49	google	cool-tshirts-search	169

Last-Touch Attribution & Campaigns

Last-touch refers to what exposure is attributed to the user returning to CTS. To break it down further, how many last-touches is each campaign responsible for?

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,  
         pv.page_name  
  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-Attribution, Con't

We can see that the email with the weekly newsletter and the FB retargeting ad brought the most people back to the site.

Query Results		
lt_attr.utm_source	lt_attr.utm_campaign	COUNT(*)
email	weekly-newsletter	447
facebook	retargeting-ad	443
email	retargeting-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

Last-Touch & Purchases

Further, we can determine how many distinct users made a purchase. That number is 361.

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

Query Results				
user_id	last_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
99933	2018-01-26 06:18:39	email	weekly-newsletter	447
99928	2018-01-24 05:26:09	facebook	retargetting-ad	443
99990	2018-01-16 11:35:09	email	retargetting-campaign	245
99589	2018-01-15 04:55:43	nytimes	getting-to-know-cool-tshirts	232
99765	2018-01-04 05:59:47	buzzfeed	ten-crazy-cool-tshirts-facts	190
99838	2018-01-02 07:40:34	medium	interview-with-cool-tshirts-founder	184
98840	2018-01-10 04:58:48	google	paid-search	178
99344	2018-01-18 21:36:32	google	cool-tshirts-search	60
COUNT(DISTINCT user_id)				361

Campaigns and Purchases

We can also determine how many last touches on the *Purchase* page each campaign is responsible for →

Query Results

last_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
2018-01-26 06:18:39	email	weekly-newsletter	115
2018-01-06 09:41:19	facebook	retargetting-ad	113
2018-01-24 09:00:58	email	retargetting-campaign	54
2018-01-19 16:37:58	google	paid-search	52
2018-01-16 15:15:29	nytimes	getting-to-know-cool-tshirts	9
2018-01-15 04:17:36	buzzfeed	ten-crazy-cool-tshirts-facts	9
2018-01-10 18:20:21	medium	interview-with-cool-tshirts-founder	7
2018-01-18 00:25:00	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)  
SELECT lt.user_id,  
       lt.last_touch_at,  
       pv.utm_source,  
       pv.utm_campaign,  
       COUNT(utm_campaign)  
FROM last_touch lt  
JOIN page_visits pv  
  ON lt.user_id = pv.user_id  
 AND lt.last_touch_at = pv.timestamp  
GROUP BY pv.utm_campaign  
ORDER BY 5 DESC;
```

Analysis & Optimizing the Campaign Budget

So, even though the interview featured on Medium brought the most people to the site (622), it was ultimately only responsible for 7 purchases & the NYTimes article (612 users to the site) resulted in 9 purchases. Despite that, I think both campaigns seem valuable for bringing traffic to the site, but, it seems as if the **email weekly newsletter** and the **retargeting ad on FB** (as well as the retargeting email campaign) resulted in the most purchases on CTS (**115, 113 & 54** respectively). I would continue to have articles and features about CTS to expand brand awareness along with funding the email campaigns and retargeting ads on FB. We want to look at the entire user journey and not solely first & last touch attribution. If we focus on the whole, we can increase purchases going forward.

Query Results			
last_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
2018-01-26 06:18:39	email	weekly-newsletter	115
2018-01-06 09:41:19	facebook	retargeting-ad	113
2018-01-24 09:00:58	email	retargeting-campaign	54
2018-01-19 16:37:58	google	paid-search	52
2018-01-16 15:15:29	nytimes	getting-to-know-cool-tshirts	9
2018-01-15 04:17:36	buzzfeed	ten-crazy-cool-tshirts-facts	9
2018-01-10 18:20:21	medium	interview-with-cool-tshirts-founder	7
2018-01-18 00:25:00	google	cool-tshirts-search	7