



Attribution Queries

Learn SQL from Scratch

Kendrick Basso

January 14th, 2019

What's Ahead

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

1. Get Familiar with CoolTShirts

1.1 How many campaigns and sources does CoolTShirts use?

CoolTShirts is an apparel company who employed several marketing campaigns to attract users to its website. The data below was taken by querying table *page_visits* where users' movement through the CoolTShirts website was tracked using several parameters. DISTINCT ensures no duplicates are included in the count.

We find eight campaigns and eight sources by querying:

```
SELECT COUNT(DISTINCT utm_campaign)
FROM page_visits;
```

```
SELECT COUNT(DISTINCT utm_source)
FROM page_visits;
```

COUNT(DISTINCT utm_campaign)

8

COUNT(DISTINCT utm_source)

6

1.2 How are the campaigns and sources related?

- *utm_campaign* is the marketing campaign or the method of bringing users' awareness to the company, which could be an advertisement, article, or search suggestion.
- *utm_source* is the mode or touchpoint from which the marketing campaign was delivered to the user, which could be a website or email.

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

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

1.3 Website Pages

The CoolTShirts website has four pages taken from the parameter *page_name*. The pages are numbered 1-4 to show the progress of a user's purchase.

```
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?

2.2 How many first touches is each campaign responsible for?

To find the first touch or initial time a user arrived on the website, we look at their first timestamp by selecting MIN(timestamp). A WITH clause creates a temporary subquery labeled *first_touch* with the minimum timestamp as a parameter. Joining *first_touch* with *page_visits*, matching the two subqueries' user ID's and timestamps, and selecting the utm_source, utm_campaign, and COUNT(utm_campaign) the number of first touches for each campaign is displayed. The GROUP BY utm_campaign categorizes the result list by utm_campaign. ORDER BY 5 DESC arranges the list by the fifth column selected in descending order.

```
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) AS  
       number_of_first_touches  
FROM first_touch ft  
JOIN page_visits pv  
    ON ft.user_id = pv.user_id  
    AND ft.first_touch_at = pv.timestamp  
GROUP BY utm_campaign  
ORDER BY 5 DESC;
```


2.3 How many first touches is each campaign responsible for?

The number_of_first_touches column shows how many first touches each campaign is responsible for:

user_id	first_touch_at	utm_source	utm_campaign	number_of_first_touches
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

2.4 How many last touches is each campaign responsible for?

To find the last touch or most recent time a user was on the website, we look at their last timestamp by selecting `MAX(timestamp)`. The rest of the query is similar to how we answered the number of first touches. A `WITH` clause creates a temporary subquery labeled *last_touch* with the maximum timestamp as a parameter. Joining *last_touch* with *page_visits*, matching the two subqueries' user ID's and timestamps, and selecting the `utm_source`, `utm_campaign`, and `COUNT(utm_campaign)` the number of last touches for each campaign is displayed. The `GROUP BY utm_campaign` categorizes the result list by `utm_campaign`. `ORDER BY 5 DESC` arranges the list by the fifth column selected in descending order.

```
WITH last_touch AS (  
    SELECT user_id,  
           MAX(timestamp) as last_touch_at  
    FROM page_visits  
    GROUP BY user_id)  
SELECT lt.user_id,  
       lt.last_touch_at,  
       pv.utm_source,  
       pv.utm_campaign,  
       COUNT(utm_campaign) AS  
       number_of_last_touches  
FROM last_touch lt  
JOIN page_visits pv  
    ON lt.user_id = pv.user_id  
    AND lt.last_touch_at = pv.timestamp  
GROUP BY utm_campaign  
ORDER BY 5 DESC;
```

2.5 How many last touches is each campaign responsible for?

The number_of_last_touches column shows how many last touches each campaign is responsible for:

user_id	last_touch_at	utm_source	utm_campaign	number_of_last_touches
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

2.6 How many visitors make a purchase?

The number of visitors who make a purchase is found by counting the number of distinct users who visited the purchase page.

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

number_of_visitors_who_purchased
361

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

Adding the clause WHERE page_name = '4 - purchase' to our previous last_touch subquery will make the campaign count specific to the purchase page

```
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) AS number_of_last_touches  
FROM last_touch lt  
JOIN page_visits pv  
  ON lt.user_id = pv.user_id  
  AND lt.last_touch_at = pv.timestamp  
GROUP BY utm_campaign  
ORDER BY 5 DESC;
```

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

The table below shows the number of last touches on the purchase page each campaign is responsible for.

user_id	last_touch_at	utm_source	utm_campaign	number_of_last_touches
99933	2018-01-26 06:18:39	email	weekly-newsletter	115
99897	2018-01-06 09:41:19	facebook	retargetting-ad	113
99285	2018-01-24 09:00:58	email	retargetting-campaign	54
94567	2018-01-19 16:37:58	google	paid-search	52
92172	2018-01-16 15:15:29	nytimes	getting-to-know-cool-tshirts	9
98651	2018-01-15 04:17:36	buzzfeed	ten-crazy-cool-tshirts-facts	9
83547	2018-01-10 18:20:21	medium	interview-with-cool-tshirts-founder	7
95650	2018-01-18 00:25:00	google	cool-tshirts-search	2

2.8 What is the typical user journey?

From the number of first touches, a user typically discovers the CoolTShirts website through an article, but seldom make a purchase. The number of last touches on the purchase page attributed to these articles is low– in the single digits. It is far more common that a user makes a purchase after revisiting the website through an email blast, retargeting ad, retargeting campaign, or paid search.

The weekly newsletter and retargeting ad were the most successful campaigns in having users make a purchase returning the highest count of last touches. The retargeting campaign and paid search brought more users to the purchase page than the news articles; however, the articles brought more attention to the website.

utm_campaign	number_of_first_touches	utm_campaign	number_of_last_touches
interview-with-cool-tshirts-founder	622	weekly-newsletter	115
getting-to-know-cool-tshirts	612	retargeting-ad	113
ten-crazy-cool-tshirts-facts	576	retargeting-campaign	54
cool-tshirts-search	169	paid-search	52
		getting-to-know-cool-tshirts	9
		ten-crazy-cool-tshirts-facts	9
		interview-with-cool-tshirts-founder	7
		cool-tshirts-search	2

3. Optimize the Campaign Budget

3.1 Optimize the campaign budget

Given that the typical user learns of CoolTShirts from an article and is brought back to the website at another time to make a purchase, the company should reinvest in three campaigns with the highest first touch counts to maximize awareness of their brand and two campaigns with the highest last touch counts on the purchase page to generate profit. Based on the numbers below, they should work with campaign sources medium who hosted the interview-with-cool-tshirts-founder article, nytimes who hosted the getting-to-know-cool-tshirts article, and buzzfeed who hosted the ten-crazy-cooltshirts-facts. CoolTShirts should also reinvest in their weekly-newsletter email blast and the retargeting-ad from facebook. Re-investing in these five campaigns based on their first and last touches would be the best use of their campaign budget.

utm_campaign	number_of_first_touches
interview-with-cool-tshirts-founder	622
getting-to-know-cool-tshirts	612
ten-crazy-cool-tshirts-facts	576

utm_campaign	number_of_last_touches
weekly-newsletter	115
retargeting-ad	113