



Attribution Queries

Learn SQL from Scratch

Marco Ruggiero

11-10-2018

Example Table of Contents

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

1. Get Familiar With Cool T-Shirts

1.1 & 1.2 Campaigns, Sources & Pages

The company has 8 distinct campaigns, & 6 distinct sources as demonstrated below. They have the below pages on their website. The associated SQL is;

```
select count (distinct utm_campaign) from page_visits;  
select count (distinct utm_source) from page_visits;  
select distinct utm_campaign, utm_source from page_visits order by utm_source asc;  
  
select distinct page_name from page_visits;
```

utm_campaign	utm_source
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

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

2. What Is The User Journey?

2.1 First Touches

The first touch figures are given below, with the 'interview with cool t-shirts founder' campaign the most popular at 622. The SQL used to derive these figures can be seen on the right.

ft_attr.utm_source	ft_attr.utm_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

```
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,
       ft_attr.utm_campaign,
       COUNT(*)
FROM ft_attr
GROUP BY 2
ORDER BY 3 DESC;
```

2.2 Last Touches

The last touch figures are given below, with the SQL used found to the right. Here we can see the most popular campaigns are the weekly newsletter through email, & the re-targeting advert campaigns through Facebook.

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

```
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,
       lt_attr.utm_campaign,
       COUNT(*)
FROM lt_attr
GROUP BY 2
ORDER BY 3 DESC;
```

2.3 Last Touches as Purchases

These figures show purchases made, with the previous SQL modified through the insertion of a 'where' clause to only return page_name equal to '4 - purchase'; confirming a purchase had been made. Here we see the top two results continue our findings from the previous query.

lt_attr.utm_source	lt_attr.utm_campaign	COUNT(*)
email	weekly-newsletter	115
facebook	retargetting-ad	113
email	retargetting-campaign	54
google	paid-search	52
nytimes	getting-to-know-cool-tshirts	9
buzzfeed	ten-crazy-cool-tshirts-facts	9
medium	interview-with-cool-tshirts-founder	7
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
),
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,
       lt_attr.utm_campaign,
       COUNT(*)
FROM lt_attr
GROUP BY 2
ORDER BY 3 DESC;
```


2.4 Typical User Journey

By changing the value of the 'where' clause in the first_touch or last_touch queries listed previously to that of the values we know from 1.2, we can track when a user stopped or completed their journey to Cool T-Shirts online shop, or which route they took. To the right, we have the example of those whose last touch was on '3 - checkout'.

From the data we have, it would seem most users are introduced to the company via one of the articles about the company, before ultimately completing any purchases through the weekly newsletter or re-targetting ads via Facebook.

The SQL here is the same as in the first_touch or last_touch queries, modifying the value of the 'where' clause to the values in page_name.

lt_attr.utm_source	lt_attr.utm_campaign	COUNT(*)
email	weekly-newsletter	450
facebook	retargetting-ad	445
email	retargetting-campaign	246
google	paid-search	179
nytimes	getting-to-know-cool-tshirts	41
buzzfeed	ten-crazy-cool-tshirts-facts	32
medium	interview-with-cool-tshirts-founder	31
google	cool-tshirts-search	7

3. Optimize The Campaign Budget

3.1 Strategy Optimization

Based upon the first & last touch results we see from the queries in 2.1, 2.2 & 2.3 (latter truncated in screenshot here), I would advise Cool T-Shirts to reinvest into the following;

- 1) interview-with-cool-tshirts-founder via medium
- 2) getting-to-know-cool-tshirts via nytimes
- 3) ten-crazy-coool-tshirts-facts via buzzfeed

...to increase general awareness of the company.
And then;

- 4) weekly-newsletter via email
- 5) retargeting-ad via facebook

...to increase sales, as these methods have shown to be considerably more effective than others in sale conversion.

ft_attr.utm_source	ft_attr.utm_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

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

lt_attr.utm_source	lt_attr.utm_campaign	COUNT(*)
email	weekly-newsletter	115
facebook	retargeting-ad	113
email	retargeting-campaign	54
google	paid-search	52
nytimes	getting-to-know-cool-tshirts	9