Attribution Queries Slide

Learn SQL from Scratch
Chen Yang
01/15/2019

Example Table of Contents

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 and how are they related? Be sure to explain the difference between utm_campaign and utm_source.

8 campaigned used **6** sources: email and google were used twice; nytimes, buzzfeed, facebook, and medium were used once. **utm_campaign** identifies the specific ad or email blast, while **utm_source** Identifies which site sent the traffic.

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

```
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;
```

1.2 What pages are on their website?

There are 4 pages on their website: landing_page, shipping_cart, checkout, and 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?

2.1 How many first touches is each campaign responsible for?

First touches can be attributed to 4 campaigns.
Cool-tshirts-search caused 169 first touches;
Getting-to-know-cool-tshirts caused 612 first touches;
Interview-with-cool-tshirts-founder caused 622 first touches;
Ten-crazy-cool-tshirts-facts caused 576 first touches.

```
WITH first_touch AS (
  SELECT user_id,
    MIN(timestamp) AS 'first_touch_at'
  FROM page_visits
  GROUP BY user_id
)
SELECT pv.utm_campaign,
       COUNT(*)
FROM first_touch AS ft
JOIN page_visits AS pv
  ON ft.user_id = pv.user_id
  AND ft.first_touch_at = pv.timestamp
GROUP BY 1;
```

utm_campaign	COUNT(*)
cool-tshirts-search	169
getting-to-know-cool-tshirts	612
interview-with-cool-tshirts-founder	622
ten-crazy-cool-tshirts-facts	576

2.2 How many last touches is each campaign responsible for?

Last touches can be attributed to all 8 campaigns. By sorting them, we can find weekly-newsletter has most last touches with 447, and cool-tshirts-search has least last touches with 60. Detail is shown below.

utm_campaign	COUNT(*)
weekly-newsletter	447
retargetting-ad	443
retargetting-campaign	245
getting-to-know-cool-tshirts	232
ten-crazy-cool-tshirts-facts	190
interview-with-cool-tshirts-founder	184
paid-search	178
cool-tshirts-search	60

```
WITH last_touch AS (
  SELECT user_id,
    MAX(timestamp) AS 'last_touch_at'
  FROM page_visits
  GROUP BY user_id
)
SELECT pv.utm_campaign,
       COUNT(*)
FROM last_touch AS lt
JOIN page_visits AS pv
  ON lt.user_id = pv.user_id
  AND lt.last_touch_at = pv.timestamp
GROUP BY 1
ORDER BY 2 DESC;
```

2.3 How many visitors make a purchase?

There are 361 visitors made a purchase.

COUNT(DISTINCT user_id)
361

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

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

Weekly-newsletter is responsible for the most last touches on purchase page, which is 115. Cool-tshirts-search is responsible for the least last touches on purchase page, which is just 2. Other campaigns are shown below.

utm_campaign	COUNT(*)
weekly-newsletter	115
retargetting-ad	113
retargetting-campaign	54
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

```
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 pv.utm_campaign,
       COUNT(*)
FROM last_touch AS lt
JOIN page_visits AS pv
  ON lt.user_id = pv.user_id
  AND lt.last_touch_at = pv.timestamp
GROUP BY 1
ORDER BY 2 DESC;
```

2.5 What is the typical user journey?

The typical user journey is: users usually are directed to CoolTShirts' website by interview-with-cool-tshirts-founder, getting-to-know-cool-tshirts, and ten-crazy-cool-tshirts-facts campaigns.
Weekly-newsletter and retargetting-ad are the two campaigns that draw most customers back to their website and make the final purchase.

3. Optimize the campaign budget

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

interview-with-cool-tshirts-founder
getting-to-know-cool-tshirts
ten-crazy-cool-tshirts-facts
weekly-newsletter
retargetting-ad

I think CoolTShirts should invest above 5 campaigns.
The reason is they need campaigns to attract enough first touches, while from the table of first touches (question 2.1) , the first 3 campaigns did pretty good on introducing products to the customers. They also need campaigns to trigger customers making the final purchase. From the result of question 2.4, we could know that weekly-newsletter and retargeting-ad did way better than other campaigns, so investing those two campaigns will be very beneficial.

utm_campaign	COUNT(*)
interview-with-cool-tshirts-founder	622
getting-to-know-cool-tshirts	612
ten-crazy-cool-tshirts-facts	576
cool-tshirts-search	169

utm_campaign	COUNT(*)
weekly-newsletter	115
retargetting-ad	113
retargetting-campaign	54
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