



Attribution Queries

Analyze Data with SQL

Denis Lazuk

2023-03-13

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 Campaigns and sources of CoolTShirts

How many campaigns and sources does CoolTShirts use? Which source is used for each campaign?

- There are 8 campaigns used by CoolTShirts;
- There are 6 sources used by CoolTShirts;
- The relation between campaign and source is given in the table below.

utm_campaign	utm_source
cool-tshirts-search	google
getting-to-know-cool-tshirts	nytimes
interview-with-cool-tshirts-founder	medium
paid-search	google
retargeting-ad	facebook
retargeting-campaign	email
ten-crazy-cool-tshirts-facts	buzzfeed
weekly-newsletter	email

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

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

```
SELECT utm_campaign,
utm_source
FROM page_visits
GROUP BY 1;
```

1.2 Pages of the CoolTShirts website

What pages are on the CoolTShirts website?

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

```
SELECT DISTINCT (page_name)
FROM page_visits;
```

2. What is the user journey?

2.1 How many first touches is each campaign responsible for?

There are 4 pairs of campaigns and sources used responsible for first touches. The actual numbers are given in the table below.

utm_campaign	utm_source	Number of first touches
interview-with-cool-tshirts-founder	medium	622
getting-to-know-cool-tshirts	nytimes	612
ten-crazy-cool-tshirts-facts	buzzfeed	576
cool-tshirts-search	google	169

```
WITH first_touch AS (  
    SELECT user_id,  
           MIN(timestamp) AS first_touch_at  
    FROM page_visits  
    GROUP BY user_id),  
result AS (  
    SELECT ft.user_id,  
           ft.first_touch_at,  
           pv.utm_source,  
           pv.utm_campaign  
    FROM first_touch AS ft  
    JOIN page_visits AS pv  
      ON ft.user_id = pv.user_id  
     AND ft.first_touch_at = pv.timestamp  
    )  
SELECT utm_campaign,  
       utm_source,  
       COUNT (*) AS 'Number of first touches'  
FROM result  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

2.2 How many last touches is each campaign responsible for?

There are 8 pairs of campaigns and sources used responsible for last touches. The actual numbers are given in the table below.

utm_campaign	utm_source	Number of last touches
weekly-newsletter	email	447
retargetting-ad	facebook	443
retargetting-campaign	email	245
getting-to-know-cool-tshirts	nytimes	232
ten-crazy-cool-tshirts-facts	buzzfeed	190
interview-with-cool-tshirts-founder	medium	184
paid-search	google	178
cool-tshirts-search	google	60

```
WITH last_touch AS (  
    SELECT user_id,  
           MAX(timestamp) AS last_touch_at  
    FROM page_visits  
    GROUP BY user_id),  
result AS (  
    SELECT lt.user_id,  
           lt.last_touch_at,  
           pv.utm_source,  
           pv.utm_campaign  
    FROM last_touch AS lt  
    JOIN page_visits AS pv  
      ON lt.user_id = pv.user_id  
      AND lt.last_touch_at = pv.timestamp  
    )  
SELECT utm_campaign,  
       utm_source,  
       COUNT (*) AS 'Number of last touches'  
FROM result  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```


2.3 How many visitors make a purchase?

There are 361 distinct users who made a purchase.

Distinct Purchases
361

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

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

The following pairs of campaigns and sources are responsible for last touches on the purchase page. The actual numbers are given in the table below.

utm_campaign	utm_source	Number of last touches
weekly-newsletter	email	114
retargetting-ad	facebook	112
retargetting-campaign	email	53
paid-search	google	52
getting-to-know-cool-tshirts	nytimes	9
ten-crazy-cool-tshirts-facts	buzzfeed	9
interview-with-cool-tshirts-founder	medium	7
cool-tshirts-search	google	2

```
WITH last_touch AS (  
    SELECT user_id,  
           MAX(timestamp) AS last_touch_at  
    FROM page_visits  
    GROUP BY user_id),  
result AS (  
    SELECT lt.user_id,  
           lt.last_touch_at,  
           pv.utm_source,  
           pv.utm_campaign  
    FROM last_touch AS lt  
    JOIN page_visits AS pv  
      ON lt.user_id = pv.user_id  
     AND lt.last_touch_at = pv.timestamp  
   WHERE page_name = '4 - purchase'  
)  
SELECT utm_campaign,  
       utm_source,  
       COUNT (*) AS 'Number of last touches'  
FROM result  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

3. Optimize the campaign budget

3 Findings for budget optimization

CoolTShirts can re-invest in 5 campaigns. Given your findings in the project, which should they pick and why?

They should pick **interview-with-cool-tshirts-founder**, **getting-to-know-cool-tshirts**, **ten-crazy-cool-tshirts-facts** as they show best numbers in customers attraction (see Table 1) and **weekly-newsletter**, **retargetting-ad** because these two show the best last touch effectiveness (see Table 2) .

Table 1

utm_campaign	utm_source	Number of first touches
interview-with-cool-tshirts-founder	medium	622
getting-to-know-cool-tshirts	nytimes	612
ten-crazy-cool-tshirts-facts	buzzfeed	576

Table 2

utm_campaign	utm_source	Number of last touches
weekly-newsletter	email	447
retargetting-ad	facebook	443