codecademy

CoolTShirts Attribution

Learn SQL from Scratch
Capstone Project - Attribution
Tian Yun (Amy) Liu
July 2018

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 Campaign and Sources

- CoolTShirts uses 8 campaigns
- There are 6 sources
- Table on the right shows which source is used for each campaign

```
Query:
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;
```

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

1.2 Website pages

- There are 4 distinct page names on the CoolTShirt website:
  - Landing_page
  - Shopping_cart
  - Checkout
  - Purchase
- See table on the right

```
Query:
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?

- 4 campaigns are responsible for all the first touches. Table on the left shows how many first touches each of them is responsible for. Medium's "interview with cool tshirts founder "seems to be responsible for the most first touches, 622 of them!

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

```
Query:

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 1, 2
ORDER BY 3 desc;
```

2.2 How many last touches is each campaign responsible for?

- Table on the next slide shows how many last touches each campaign is responsible for. Weekly newsletter via email and Retargetting ad via Facebook seem to be effective as they are responsible for the most last touches

```
Query:
 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 1,2
 ORDER by 3 desc;
```

2.2 How many last touches is each campaign responsible for? (Cont'd)

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

2.3 How many visitors make a purchase

- 361 visitors made a purchase

```
Query:
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?

- Table on the next slide shows how many last touches on the purchase page each campaign is responsible for. Again, Weekly newsletter via email and Retargetting ad via Facebook are at the top. Google paid search, interestingly, didn't yield as many last touches in total than some other campaigns but did yield more purchases compared to i.e. nytimes' campaign

```
Query:
 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 1,2
ORDER by 3 desc;
```

2.4 How many last touches is each campaign responsible for? (Cont'd)

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
buzzfeed	ten-crazy-cool-tshirts-facts	9
nytimes	getting-to-know-cool-tshirts	9
medium	interview-with-cool-tshirts-founder	7
google	cool-tshirts-search	2

2.5 Typical user journey

Based on the data we have in earlier slides, users usually reach the site through a campaign (NYTimes, Medium, Buzzfeed articles, etc) and have their first visits. A lot of them end up leaving the site before getting to the "add to cart" page or the "purchase" page. For the ones that come back and/or add a product to the cart, most of them come back a few days later through other campaigns (facebook and email retargetting campaigns, etc). I used the 2 queries below to double check my conclusion. I also saw most of the users made it through more than just the landing page.

```
Query:
SELECT user_id, page_name, timestamp
FROM page_visits
ORDER BY user_id;

SELECT user_id, count(DISTINCT page_name)
FROM page_visits
GROUP BY user_id
ORDER BY 2 desc;
```

3. Optimize the campaign budget

Which 5 campaigns should CoolTshirts re-invest in?

I recommend CoolTshirts to re-invest in the following 5 campaigns, with the goal of attracting users to visit the site and make purchases:

1. Weekly-newsletter via Email - high last touch with purchase
2. Retargetting-ad via Facebook - high last touch with purchase
3. Retargetting-campaign via Email - high last touch with purchase
4. Getting-to-know-cool-tshirts via NYTimes - high first touch and decent last touch with some purchases
5. Interview-with-cool-tshirts-founder via Medium - high first touch

Since users tend to leave the site after first visit but come back to purchase via other campaigns, using the NYTimes and Medium campaigns to attract users first and retarget those who left the site later using email and facebook campaigns seem to be the way to go