



Capstone: Attribution

Learn SQL from Scratch

Kate Kim

August 26, 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 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`.

As you can see from the tables below, CoolTShirts use 6 sources and 8 campaigns.

The difference between `utm_campaign` and `utm_source`:

- **`utm_campaign`** identifies the specific advertising or emails the users received resulting in the users visiting the website
- **`utm_source`** identifies the channels led the users to CoolTShirts.

```
select count(distinct utm_source)
from page_visits;
```

```
select count(distinct utm_campaign)
from page_visits;
```

```
select distinct utm_source, utm_campaign
from page_visits;
```

count(distinct utm_source)
6

count(distinct utm_campaign)
8

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

1.2 What Pages are on their website?

There are 4 pages on the website as per the table below.

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 What is the user journey? – How many first touches is each campaign responsible for?

The table below shows the campaigns and the volume of first touches each campaign is responsible for.

utm_campaign	count(ft.first_touch_at)
cool-tshirts-search	169
getting-to-know-cool-tshirts	612
interview-with-cool-tshirts-founder	622
ten-crazy-cool-tshirts-facts	576

```
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(ft.first_touch_at)  
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;
```

2.2 What is the user journey? – How many last touches is each campaign responsible for?

The table below shows the campaigns and the volume of last touches each campaign is responsible for.

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

```
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(lt.last_touch_at)  
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;
```


2.3 What is the user journey? – How many visitors make a purchase?

In total, 361 visitors make a purchase.

count(distinct user_id)
361

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

```
select distinct page_name  
from page_visits;
```

```
select count(distinct user_id)  
from page_visits  
where page_name = '4 - purchase';
```

2.4 What is the user journey? – How many last touches on the purchase page is each campaign responsible for?

The first column below is the campaign, and the next column is the volume of last touches led to the purchase page per corresponding campaign

utm_campaign	count(lt.last_touch_at)
cool-tshirts-search	2
getting-to-know-cool-tshirts	9
interview-with-cool-tshirts-founder	7
paid-search	52
retargeting-ad	112
retargeting-campaign	53
ten-crazy-cool-tshirts-facts	9
weekly-newsletter	114

```
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(lt.last_touch_at)  
FROM last_touch lt  
JOIN page_visits pv  
  ON lt.user_id = pv.user_id  
AND lt.last_touch_at = pv.timestamp  
where pv.page_name = '4 - purchase'  
group by pv.utm_campaign;
```

2.5 What is the user journey? – What is the typical user journey?

1. Typically, users click a link inside the articles through the famous sources such as medium, nytimes, buzzfeed or google. The link will lead to CoolTShirt website. The most popular source is medium with the campaign called “interview-with-cool-tshirts-founder”, boasting the highest number of first website (CoolTShirts) visit through the campaign.
2. The users however do not purchase the items straight away through that link.
They look around and go to other websites and some might come back later for further browsing or purchase.
3. The most effective campaigns used are “weekly-newsletter” via email and “retargeting-ad” via facebook.
The users then decide to purchase the products they have forgotten about or thought about for awhile.

2.5 What is the user journey? – What is the typical user journey? – continued..

User_id	utm_campaign	utm_source	page_name	count(lt.last_touch_at)
99933	weekly-newsletter	email	4 - purchase	447
99928	retargeting-ad	facebook	3 - checkout	443
99990	retargeting-campaign	email	3 - checkout	245
99589	getting-to-know-cool-tshirts	nytimes	2 - shopping_cart	232
99765	ten-crazy-cool-tshirts-facts	buzzfeed	2 - shopping_cart	190
99838	interview-with-cool-tshirts-founder	medium	3 - checkout	184
98840	paid-search	google	3 - checkout	178
99344	cool-tshirts-search	google	2 - shopping_cart	60

user_id	utm_campaign	utm_source	page_name	count(ft.first_touch_at)
99990	interview-with-cool-tshirts-founder	medium	1 - landing_page	622
99933	getting-to-know-cool-tshirts	nytimes	1 - landing_page	612
99765	ten-crazy-cool-tshirts-facts	buzzfeed	1 - landing_page	576
99684	cool-tshirts-search	google	1 - landing_page	169

2.5 What is the user journey? – What is the typical user journey? – continued..

```
WITH last_touch AS (  
  SELECT user_id, Max(timestamp) as last_touch_at  
  FROM page_visits  
  GROUP BY user_id  
)  
SELECT pv.user_id, pv.utm_campaign, pv.utm_source, pv.page_name, count(lt.last_touch_at)  
FROM last_touch lt JOIN page_visits pv  
ON lt.user_id = pv.user_id  
AND lt.last_touch_at = pv.timestamp  
group by pv.utm_campaign  
order by count(lt.last_touch_at) desc;
```

```
WITH first_touch AS (  
  SELECT user_id, min(timestamp) as first_touch_at  
  FROM page_visits  
  GROUP BY user_id  
)  
SELECT pv.user_id, pv.utm_campaign, pv.utm_source, pv.page_name, count(ft.first_touch_at)  
FROM first_touch ft  
JOIN page_visits pv  
ON ft.user_id = pv.user_id  
AND ft.first_touch_at = pv.timestamp  
group by pv.utm_campaign  
order by count(ft.first_touch_at) desc;
```

3. Optimize the campaign budget

3. Optimize the campaign budget - CoolTShirts can re-invest in 5 campaigns. Which should they pick and why?

- interview-with-cool-tshirts-founder (1st touch)
- getting-to-know-cool-tshirts (1st touch)
- weekly-newsletter (last touch)
- retargeting-ad (last touch)
- retargeting-campaign (last touch)

Why?

The first two selections clearly show the highest clicks that led to the exposure of the website & products to the customers. They are working in terms of attracting people onto the website. The last three campaigns are all re-targeting campaigns. The the purchase (or close to purchase) trend for this company is highly attributable to retargeting marketing via email or facebook.



Capstone: Attribution

End of the slide...

Thank you.