Learning SQL from Scratch: First and Last Touch Attribution with CoolTShirts.com

Capstone Project 2 - Harrison Dahl

07/02/2018

Objective

Reason for analysis:

1. Get familiar with the company.

- 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.
- What pages are on their website?

2. What is the user journey?

- How many first touches is each campaign responsible for?
- How many last touches is each campaign responsible for?
- How many visitors make a purchase?
- How many last touches *on the purchase page* is each campaign responsible for?
- What is the typical user journey?

3. Optimize the campaign budget

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

Query Results

Number of Campaigns	
8	

Number of Sources	
6	

Campaign Name	Source Name
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

Page Name
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

Getting to know CoolTShirts from initial .csv data

The data indicates that CoolTShirts.com has 8 marketing campaigns that draw clicks from 6 utm sources

These 6 utm sources range from the NY Times to Buzzfeed with multiple campaigns targeted at Google and Email

Using a simple select distinct query, we find that CoolTShirts.com has 4 pages on their website. This seems like a very small number but these may be categories enveloping many smaller web pages

User Journey

First Touches:

We can use the MIN(timestamp) clause to find the first touches for each user_id

We find that there are 4 campaign sources for first touches, which is obviously only ½ of the campaigns the company is currently running

We can infer that the 4 campaigns not generating first touches might be campaigns that were not follow-up campaigns

```sql
1   WITH first_touch AS (
2       SELECT user_id,
3           MIN(timestamp) as first_touch_at
4       FROM page_visits
5       GROUP BY user_id),
6
7   ft_attr as (
8     select ft.user_id,
9           ft.first_touch_at,
10          pv.utm_source,
11          pv.utm_campaign
12    from first_touch ft
13    join page_visits pv
14      on ft.user_id =pv.user_id
15      and ft.first_touch_at = pv.timestamp)
16
17  SELECT ft_attr.utm_source as Source,
18         ft_attr.utm_campaign as Campaign,
19         count(*) as count
20  FROM ft_attr
21  group by 1, 2
22  order by 3 desc;
```

Query Results		
Source	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

User Journey

Last Touches:

We can use the MAX(timestamp) clause this time to find the last touches for each user_id

We find that the first touch target campaigns have lowered in count

We see that our hypothesis was correct and that the majority of these touches come from retargetting campaigns

```
1   WITH last_touch AS (
2       SELECT user_id,
3           MAX(timestamp) as last_touch_at
4       FROM page_visits
5       GROUP BY user_id),
6
7   ft_attr as (
8     select lt.user_id,
9           lt.last_touch_at,
10          pv.utm_source,
11          pv.utm_campaign
12    from last_touch lt
13    join page_visits pv
14      on lt.user_id =pv.user_id
15      and lt.last_touch_at = pv.timestamp)
16
17  SELECT ft_attr.utm_source as Source,
18         ft_attr.utm_campaign as Campaign,
19         count(*) as Count
20  FROM ft_attr
21  group by 1, 2
22  order by 3 desc;
```

Query Results		
Source	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

User Journey

Number of Visitors that Make a Purchase:

We can use a simple select count(distinct user_id) clause to find that we have 361 customers that purchase out of the 1979 total user_id count

This is an 18.24% conversion rate to purchase which is incredibly high for a standard website

```
17    select count(distinct user_id) as 'Visitor Purchase Count'
18    from page_visits where page_name = '4 - purchase';
```

Query Results
Visitor Purchase Count
361

User Journey

Campaigns Responsible for Last Touch Purchases:

We can use the MAX(timestamp) clause again to find the last touches from the "4 - purhcase page.. Then we can use a select clause to categorize the utm sources and campaigns by count(*)

It looks like most last touch purchases were once again made by retargetting ads

```sql
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),

ft_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 ft_attr.utm_source as Source,
        ft_attr.utm_campaign as Campaign,
        count(*) as Count
FROM ft_attr
group by 1, 2
order by 3 desc;
```

Query Results

Source	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

The Typical User Journey

We find that 9/10 campaigns that cause first touches come from 4 sources, of which almost ⅓ of the traffic comes from an interview of their (probably charismatic) Founder/CEO

However, these initial 4 campaigns severely drop off in repeat or last touches and don't often result in purchases. 78 percent of purchases were made from the weekly newsletter or retargetting campaigns

To me, it seems like the most important thing for this company to do is to **acquire email addresses** using the best means necessary; as this translates to both making purchases from the newsletter or a retargetting campaign

Recommendation

CoolTShirts can only invest in 5 campaigns. I would recommend continuing to push articles and interviews in order to acquire emails for the newsletter and retargetting

It is hard from this data to know how many emails were captured from the first touch campaigns. I would strongly recommend finding a way to acquire this data ASAP in order to target correctly

Campaign	First Touch Count	Last Touch Count	Purchase Count	% Total First Touch	% Total Last Touch	% Total Purchases
Founder Interview	622	184	7	31%	9%	2%
NY Times Article	612	232	9	31%	12%	2%
Buzzfeed Article	576	190	9	29%	10%	2%
Web Search	169	60	2	9%	3%	1%
Newsletter	0	447	115	0%	23%	32%
Retargetting Ad	0	443	113	0%	22%	31%
Retargetting Campaign	0	245	54	0%	12%	15%
Paid Search Result	0	178	52	0%	9%	14%

Thank You