code|cademy

Capstone: Attribution
Learn SQL from Scratch
Ji Hwan Hwang
June 14th, 2018

Example Table of Contents

1. Get familiar with CoolTShirts
2. What is the user journey?
3. Optimize the campaign budget

1. Get familiar
with the company

1.1 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

-> CoolTshirts use 8 campaigns and 6 sources. In case of CoolTShirts, utm_campaign identifies the specific ad or email blast (i.e., June-21-newsletter or memorial-day-sale), and utm_source identifies which site sent the traffic (i.e., google, newsletter, or facebook_ad).

-> Data from the right queries are on the next slide.

```
-- 1.1 Get familiar with the company

select count (distinct utm_campaign) as 'count_campaign'
from page_visits;

select count (distinct utm_source) as 'count_source'
from page_visits;

select distinct utm_campaign as campaign,
utm_source as source
from page_visits;
```

1.1 Get familiar with the company

Count Campaign
8

Count Source
6

Campaign	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 Get familiar with the company

- What pages are on their website?

-> There are landing page, shopping cart, checkout, and purchase pages on the website.

Page names
1 – landing_page
2 – shopping_cart
3 – checkout
4 - purchase

```sql
-- You can put your query here

select distinct page_name as 'page_names'
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?

-> Among 8 campaigns, only 4 campaigns are responsible for first touches. Specifically, interview-with-cool-tshirts-founder has 622 touches, getting-to-know-cool-tshirts has 612 counts, ten-crazy-cool-tshirts-facts has 576 counts, and cool-tshirts-search has 169 counts.

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

```
-- How many first touches is each campaign responsible
-- for?

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 as 'Source',
ft_attr.utm_campaign as 'Campaign',
count(*)
from ft_attr
group by 1,2
order by 3 desc;
```

2.2 What is the user journey

- How many last touches is each campaign responsible for?

-> Weekly-newsletter of email has 447 counts, retargetting-ad of facebook has 443 counts, retargeting-campaign of email has 245 counts, getting-to-know-cool-tshirts of nytimes has 232 counts, ten-crazy-cool-tshirts-facts of buzzfeed has 190 counts, interview-with-cool-tshirts-founder of medium has 184 counts, paid-search of google has 178 counts, and cool-tshirts-search of google has 60 counts.

-> According to the table, weekly-newsletter campaign of email has the highest number of counts, and cool-tshirts-search has the lowest number of counts.

```
-- You can put your query here

with last_touch as(
            select user_id,
            max(timestamp) as last_touch_at
            from page_visits
            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(*)
from ft_attr
group by 1,2
order by 3 desc;
```

2.2 What is the user journey

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

2.3 What is the user journey

- How many visitors make a purchase?

-> The visitors who made a purchase are 361.
-> 361/1979 = 0.18 = 18%

Number of Visitors
361

```
-- You can put your query here

select count (distinct user_id) as 'Number of
Visitors'
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?

-> Weekly-newsletter of email has 115 counts, retargeting-ad of facebook has 113 counts, retargeting-campaign of email has 54 counts, paid-search of google has 52 counts, ten-crazy-cool-tshirts-facts of buzzfeed has 9 counts, getting-to-know-cool-tshirts of nytimes has 9 counts, interview-with-cool-tshirts-founder of medium has 7 counts, and cool-tshirts-search of google has 2 counts.

-> Among campaigns, weekly-newsletter of email has the highest number of counts, and cool-tshirts-search of google has the lowest number of counts.

```
-- You can put your query here

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(*)
from ft_attr
group by 1,2
order by 3 desc;
```
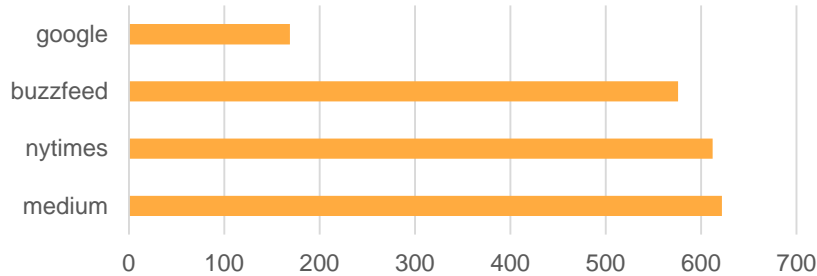
2.4 What is the user journey

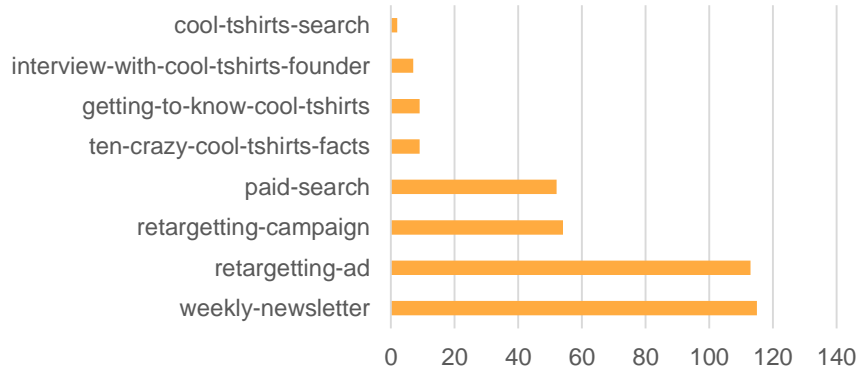| Source	Campaign	Count(*)
email	weekly-newsletter	115
facebook	retargetting-ad	113
email	retargetting-campaign	54
google	paid-search	52
buzzfeed	ten-crazy-cool-tshirts-facts	9
nytimes	getting-to-know-cool-tshirts	9
medium	interview-with-cool-tshirts-founder	7
google	cool-tshirts-search	2

2.5 What is the typical user journey

First Touches

google	(~170)
buzzfeed	(~580)
nytimes	(~615)
medium	(~620)

0   100   200   300   400   500   600   700

Last Touches

cool-tshirts-search	(~60)
paid-search	(~180)
interview-with-cool-tshirts-founder	(~185)
ten-crazy-cool-tshirts-facts	(~190)
getting-to-know-cool-tshirts	(~230)
retargetting-campaign	(~245)
retargetting-ad	(~445)
weekly-newsletter	(~450)

0   100   200   300   400   500

Purchases

cool-tshirts-search	(~3)
interview-with-cool-tshirts-founder	(~8)
getting-to-know-cool-tshirts	(~12)
ten-crazy-cool-tshirts-facts	(~12)
paid-search	(~52)
retargetting-campaign	(~55)
retargetting-ad	(~112)
weekly-newsletter	(~115)

0   20   40   60   80   100   120   140

->In case of first touches, buzzfeed, nytimes, and medium are the typical user journey.

->However, these typical user journey do not lead to the campaigns of the last touches. Specifically, retargeting-campaign, retargeting-ad, and weekly-newspaper are the different journey contrast to the first touches.

-> Paid-search, retargeting-campaign, retargeting-ad, and weekly-newsletter took account of most part of purchases of visitors.

3. Optimize the campaign budget

3.1 Optimize the campaign budget

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

Campaign	First Touches	Percentage of First Touches	Last Touches	Percentage of Last Touches	Purchases	Percentage of Purchases
getting-to-know-cool-tshirts	612	31%	232	12%	9	2%
weekly-newsletter	0	0%	447	23%	115	32%
ten-crazy-cool-tshirts-facts	576	29%	190	10%	9	2%
retargetting-campaign	0	0%	245	12%	54	15%
retargetting-ad	0	0%	443	22%	113	31%
interview-with-cool-tshirts-founder	622	31%	184	9%	7	2%
paid-search	0	0%	178	9%	52	14%
cool-tshirts-search	169	9%	60	3%	2	1%

3.1 Optimize the campaign budget

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

-> If CoolTShirts can re-invest in 5 campaigns, they should mainly focus on three campaigns, such as interview-with-cool-tshirts-founder, getting-to-know-cool-tshirts, and ten-crazy-cool-tshirts-facts. These three campaigns has the highest percentage of the first touches. They need to invest in these three campaigns more to lead the higher percentage of the first touches.

-> Next, if we look the total percentage of purchases, weekly-newsletter, retargeting-ad, and retargeting-campaign have the highest percentage of purchases. Thus, they need to re-invest in two of these campaigns to lead to increase purchases of the visitors.