



Attribution Queries

Learn SQL from Scratch

Lee Seng Yong Alvin

31/3/2019

Table of Contents

1. Get familiar with the CoolTShirts.

- 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?

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`.

There are 8 Distinct Campaigns & 6 Distinct Sources.

In 6 Distinct Source, there are 8 Distinct Campaign:

- “nytimes” has 1 Campaign (getting-to-know-cool-tshirts)
- “buzzfeed” has 1 Campaign (ten-crazy-cool-tshirts-facts)
- “facebook” has 1 Campaign (retargeting-ad)
- “medium” has 1 Campaign (interview-with-cool-tshirts-founder)
- “email” have 2 Campaign (weekly-newsletter & retargeting-campaign)
- “google” have 2 Campaign (paid-search & cool-tshirts-search)

COUNT (DISTINCT utm_campaign)

8

COUNT (DISTINCT utm_source)

6

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

```
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;
```

1.2 What pages are on their website?

We want to find all the distinct page names.

There are 4 pages in CoolTshirts Website, whereby page name are:

- **Landing_page**
- **Shopping_cart**
- **Checkout**
- **Purchase**

```
SELECT DISTINCT page_name  
from page_visits;
```

Query Results

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?

The first touches for the 4 campaigns are grouped by campaign from highest to lowest first touches:

- **Campaign #1 (Interview-with-cool-tshirts-founder): 622 first touch**
- **Campaign #2 (getting-to-know-cool-tshirts): 612 first touch**
- **Campaign #3 (ten-crazy-cool-shirts-facts): 576 first touch**
- **Campaign #4 (cool-tshirts-search): 169 first touch**

Query Results

user_id	first_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
99990	2018-01-13 23:30:09	medium	interview-with-cool-tshirts-founder	622
99933	2018-01-25 00:04:39	nytimes	getting-to-know-cool-tshirts	612
99765	2018-01-04 05:59:46	buzzfeed	ten-crazy-cool-tshirts-facts	576
99684	2018-01-13 13:20:49	google	cool-tshirts-search	169

```
WITH first_touch AS (  
  SELECT user_id,  
         MIN(timestamp) as first_touch_at  
  FROM page_visits  
  GROUP BY user_id)  
SELECT ft.user_id,  
       Ft.first_touch_at,  
       Pv.utm_source,  
       Pv.utm_campaign,  
       COUNT(utm_campaign)  
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  
ORDER BY 5 DESC;
```

2.2 How many last touches is each campaign responsible for?

The last touches for the 8 campaigns are grouped by campaign from highest to lowest last touches:

- Campaign #1 (weekly-newsletter): 447 last touch
- Campaign #2 (retargeting-ad): 443 last touch
- Campaign #3 (retargeting-campaign): 245 last touch
- Campaign #4 (getting-to-know-cool-tshirts): 232 last touch
- Campaign #5 (ten-crazy-cool-tshirts-facts): 190 last touch
- Campaign #6 (interview-with-cool-tshirts-founder): 184 last touch
- Campaign #7 (paid-search): 178 last touch
- Campaign #8 (cool-tshirts-search): 60 last touch

Query Results				
user_id	last_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
99933	2018-01-26 06:18:39	email	weekly-newsletter	447
99928	2018-01-24 05:26:09	facebook	retargeting-ad	443
99990	2018-01-16 11:35:09	email	retargeting-campaign	245
99589	2018-01-15 04:55:43	nytimes	getting-to-know-cool-tshirts	232
99765	2018-01-04 05:59:47	buzzfeed	ten-crazy-cool-tshirts-facts	190
99838	2018-01-02 07:40:34	medium	interview-with-cool-tshirts-founder	184
98840	2018-01-10 04:58:48	google	paid-search	178
99344	2018-01-18 21:36:32	google	cool-tshirts-search	60

```
WITH last_touch AS (  
  SELECT user_id,  
         MAX(timestamp) as last_touch_at  
  FROM page_visits  
  GROUP BY user_id)  
SELECT lt.user_id,  
       lt.last_touch_at,  
       pv.utm_source,  
       pv.utm_campaign,  
       COUNT(utm_campaign)  
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  
ORDER BY 5 DESC;
```


2.3 How many visitors make a purchase?

Count the distinct users who visited the page named 4 - purchase:

- **361 Distinct Visitors** make a purchase

Query Results			
user_id	last_touch_at	utm_source	utm_campaign
99933	2018-01-26 06:18:39	email	weekly-newsletter
99928	2018-01-24 05:26:09	facebook	retargetting-ad
99990	2018-01-16 11:35:09	email	retargetting-campaign
99589	2018-01-15 04:55:43	nytimes	getting-to-know-cool-tshirts
99765	2018-01-04 05:59:47	buzzfeed	ten-crazy-cool-tshirts-facts
99838	2018-01-02 07:40:34	medium	interview-with-cool-tshirts-four
98840	2018-01-10 04:58:48	google	paid-search
99344	2018-01-18 21:36:32	google	cool-tshirts-search
COUNT(DISTINCT user_id)			
361			

```
WITH last_touch AS (  
  SELECT user_id,  
         MAX(timestamp) as last_touch_at  
  FROM page_visits  
  GROUP BY user_id)  
SELECT lt.user_id,  
       lt.last_touch_at,  
       pv.utm_source,  
       pv.utm_campaign,  
       COUNT(utm_campaign)  
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  
ORDER BY 5 DESC;
```

```
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?

The last touches on purchase page (who purchased) for the 8 campaigns are grouped by campaign from highest to lowest last touches:

- **Campaign #1 (weekly-newsletter): 115 last touch**
- **Campaign #2 (retargeting-ad): 113 last touch**
- **Campaign #3 (retargeting-campaign): 54 last touch**
- **Campaign #4 (paid-search): 52 last touch**
- **Campaign #5 (getting-to-know-cool-tshirts): 9 last touch**
- **Campaign #6 (ten-crazy-cool-tshirts-facts): 9 last touch**
- **Campaign #7 (interview-with-cool-tshirts-founder): 7 last touch**
- **Campaign #8 (cool-tshirts-search): 2 last touch**

Query Results

user_id	last_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
99933	2018-01-26 06:18:39	email	weekly-newsletter	115
99897	2018-01-06 09:41:19	facebook	retargeting-ad	113
99285	2018-01-24 09:00:58	email	retargeting-campaign	54
94567	2018-01-19 16:37:58	google	paid-search	52
92172	2018-01-16 15:15:29	nytimes	getting-to-know-cool-tshirts	9
98651	2018-01-15 04:17:36	buzzfeed	ten-crazy-cool-tshirts-facts	9
83547	2018-01-10 18:20:21	medium	interview-with-cool-tshirts-founder	7
95650	2018-01-18 00:25:00	google	cool-tshirts-search	2

```
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)  
SELECT lt.user_id,  
       lt.last_touch_at,  
       pv.utm_source,  
       pv.utm_campaign,  
       COUNT(utm_campaign)  
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  
ORDER BY 5 DESC;
```

2.5 What is the typical user journey?

The typical user journey comes from these 4 campaigns (weekly-newsletter, retargeting-ad, retargeting-campaign & paid-search); which means campaigns that are repeated over a consistent basis are likely to outperform those campaigns that are not repeated.

First touches:

- Campaign (Interview-with-cool-tshirts-founder): 622 first touch
- Campaign (getting-to-know-cool-tshirts): 612 first touch
- Campaign (ten-crazy-cool-shirts-facts): 576 first touch
- Campaign (cool-tshirts-search): 169 first touch

Last touches (361 visitors made a purchase):

- Campaign (weekly-newsletter): 115 last touch = 100%
- Campaign (retargeting-ad): 113 last touch = 100%
- Campaign (retargeting-campaign): 54 last touch = 100%
- Campaign (paid-search): 52 last touch = 100%
- Campaign (getting-to-know-cool-tshirts): 9 last touch = $9/612 = 1.5\%$
- Campaign (ten-crazy-cool-tshirts-facts): 9 last touch = $9/576 = 1.6\%$
- Campaign (interview-with-cool-tshirts-founder): 7 last touch = $7/622 = 1.1\%$
- Campaign (cool-tshirts-search): 2 last touch = $2/169 = 1.2\%$

3. Optimize the campaign budget

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

CoolTShirts should re-invest in 5 campaigns below as these campaigns have are the Top 5 highest last touches & highest rate % of conversion from first to last touch (purchased):

- Campaign #1 (weekly-newsletter): 115 last touch [100% Conversion]
- Campaign #2 (retargetting-ad): 113 last touch [100% Conversion]
- Campaign #3 (retargetting-campaign): 54 last touch [100% Conversion]
- Campaign #4 (paid-search): 52 last touch [100% Conversion]
- Campaign #5 (ten-crazy-cool-tshirts-facts): 9 last touch [1.6% Conversion]