code|cademy

First and Last Touch Attribution

Learn SQL from Scratch
ALINA NICODIN
2018-08-24

Table of Contents

1. Get familiar with CoolTShirts

1.1 How many campaigns and sources does CoolTShirts use and how are they related?

- A campaign is a specific message that targets a small group of potential customers that are interested in a specific product.
- A source is an ad channel and identifies from where the potential customers came from (email, Google, Buzzfeed, Facebook, Medium, The New York Times).
- *CoolTShirts* has 8 different campaigns and 6 different sources.
- A source can run multiple campaigns, like *email* on which *weekly-newsletter* and *retargetting-campaign* are shared.
- After running the last code for every campaign (16 – 19 lines) it results that each campaign is promoted on only one source.
- In conclusion, a source can run multiple campaigns, but a source can't be run across two or more sources.

```sql
1    -- 1.1 Count campaigns and sources and find the
     relationship between them --
2
3    /* The number of distinct campaigns */
4    SELECT COUNT(DISTINCT utm_campaign) AS 'Campaign
     Count'
5    FROM page_visits;
6
7    /* The number of distinct sources */
8    SELECT COUNT(DISTINCT utm_source) AS 'Source Count'
9    FROM page_visits;
10
11   /* Relationship between campaigns and sources */
12   SELECT DISTINCT utm_campaign AS 'Campaigns',
13                   utm_source AS 'Sources'
14   FROM page_visits;
15
16   SELECT utm_campaign AS 'Campaigns',
17          utm_source AS 'Sources'
18   FROM page_visits
19   WHERE utm_campaign = 'cool-tshirts-search';
```

Campaign Count	Source Count
8	6

Campaigns	Sources
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. Get familiar with CoolTShirts

1.2 What pages are on their website?

- All the 8 campaigns are sending users to 4 different pages through the ad channels mentioned in the previous slide.
- All these pages represents the customers' journey, from visiting the website to purchasing a specific product. Note that none of them is related to *shipping conditions* or *contact* because the purpose of the marketing campaigns is to increase the traffic and purchases.

```
21    -- 1.2 Pages that are on the CoolTShirts website--
22    SELECT DISTINCT page_name AS 'Page Names'
23    FROM page_visits;
```

Page Names
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?

○ After counting the total number of first touches for each campaign, the results suggest that only half of them captured the potential customers attention.

```sql
1   -- 2.1 Count first touches per campaign and source --
2
3   /* Create a temporary table that returns the first touches for each
    user_id */
4   WITH first_touch AS (
5       SELECT user_id,
6               MIN(timestamp) AS first_touch_at
7       FROM page_visits
8       GROUP BY user_id),
9   /* Create another temporary table adding source and campaign columns by
    joining them on user_id and timestamp */
10      ft_attr AS (
11      SELECT ft.user_id,
12              ft.first_touch_at,
13              pv.utm_source,
14              pv.utm_campaign
15      FROM first_touch AS ft
16      JOIN page_visits AS pv
17        ON ft.user_id = pv.user_id
18        AND ft.first_touch_at = pv.timestamp)
19  /* Create a table that includes the total number of first touches group
    by source and campaign */
20  SELECT ft_attr.utm_source AS 'Source',
21          ft_attr.utm_campaign AS 'Campaign',
22          COUNT(*) AS 'First Touch Count'
23  FROM ft_attr
24  GROUP BY 1, 2
25  ORDER BY 3 DESC;
```

Source	Campaign	First Touch 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

5

2. What is the user journey?

2.1 How many first touches is each campaign responsible for?

○ Visitors initially discovered the *CoolTShirts* website through Medium, The New York Times, Buzzfeed and Google.

```sql
-- 2.1 Count first touches per campaign and source (in percentage terms) --

/* Create a temporary table that returns the first touches for each user_id */
WITH first_touch AS (
    /*...*/),
/* Create a temporary table that includes the total number of first touches group by
source and campaign */
    counting AS (
    SELECT ft_attr.utm_source AS source,
           ft_attr.utm_campaign AS campaign,
           COUNT(*) AS ft_count
    FROM ft_attr
    GROUP BY 1, 2
    ORDER BY 3 DESC),
/* Create a temporary table that includes the total number of first touches */
    total_counting AS (
    SELECT SUM(counting.ft_count) AS ft_total
    FROM counting),
/* Create a temporary table that is a CROSS JOIN between counting table and
total_counting table */
    cj AS (
    SELECT *
    FROM counting
    CROSS JOIN total_counting)
/* Create a table that includes the total number of first touches in percentage
terms group by source and campaign */
SELECT cj.source AS 'Source',
       cj.campaign AS 'Campaign',
       ROUND(100.0 * cj.ft_count / cj.ft_total,2) AS '% 1st touches'
FROM cj
GROUP BY 1, 2
ORDER BY 3 DESC;
```

Source	Campaign	% 1st touches
medium	interview-with-cool-tshirts-founder	31.43
nytimes	getting-to-know-cool-tshirts	30.92
buzzfeed	ten-crazy-cool-tshirts-facts	29.11
google	cool-tshirts-search	8.54

2. What is the user journey?

2.2 How many last touches is each campaign responsible for?

○ After counting the total number of last touches for each campaign, the results suggest that all of them succeed in capturing the potential customers attention for the second time.

```sql
1  -- 2.2 Count last touches per campaign and source --
2
3  /* Create a temporary table that returns the last touches for each
   user_id */
4  WITH last_touch AS (
5      SELECT user_id,
6          MAX(timestamp) AS last_touch_at
7      FROM page_visits
8      GROUP BY user_id),
9  /* Create another temporary table adding source and campaign columns by
   joining them on user_id and timestamp */
10     lt_attr AS (
11     SELECT lt.user_id,
12          lt.last_touch_at,
13          pv.utm_source,
14          pv.utm_campaign
15     FROM last_touch AS lt
16     JOIN page_visits AS pv
17       ON lt.user_id = pv.user_id
18       AND lt.last_touch_at = pv.timestamp)
19  /* Create a table that includes the total number of last touches group
   by source and campaign */
20  SELECT lt_attr.utm_source AS 'Source',
21        lt_attr.utm_campaign AS 'Campaign',
22        COUNT(*) AS 'Last Touch Count'
23  FROM lt_attr
24  GROUP BY 1, 2
25  ORDER BY 3 DESC;
```

Source	Campaign	Last Touch 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. What is the user journey?

2.2 How many last touches is each campaign responsible for?

○ Interested to note that most of the visitors are drawn back to *CoolTShirts* by the campaigns that weren't responsible for the first touches.

```
1   -- 2.2 Count last touches per campaign and source (in percentage terms) --
2
3   /* Create a temporary table that returns the last touches for each user_id */
4   WITH last_touch AS (
5       /* ... */),
6   /* Create a temporary table that includes the total number of last touches group by
    source and campaign */
7       counting AS (
8       SELECT lt_attr.utm_source AS source,
9              lt_attr.utm_campaign AS campaign,
10             COUNT(*) AS lt_count
11      FROM lt_attr
12      GROUP BY 1, 2
13      ORDER BY 3 DESC),
14  /* Create a temporary table that includes the total number of last touches */
15      total_counting AS (
16      SELECT SUM(counting.lt_count) AS lt_total
17      FROM counting),
18  /* Create a temporary table that is a CROSS JOIN between counting table and
    total_counting table */
19      cj AS (
20      SELECT *
21      FROM counting
22      CROSS JOIN total_counting)
23  /* Create a table that includes the total number of last touches in percentage
    terms group by source and campaign */
24  SELECT cj.source AS 'Source',
25         cj.campaign AS 'Campaign',
26         ROUND(100.0 * cj.lt_count / cj.lt_total,2) AS '% Last touches'
27  FROM cj
28  GROUP BY 1, 2
29  ORDER BY 3 DESC;
```

Source	Campaign	% Last touches
email	weekly-newsletter	22.59
facebook	retargetting-ad	22.39
email	retargetting-campaign	12.38
nytimes	getting-to-know-cool-tshirts	11.72
buzzfeed	ten-crazy-cool-tshirts-facts	9.6
medium	interview-with-cool-tshirts-founder	9.3
google	paid-search	8.99
google	cool-tshirts-search	3.03

2. What is the user journey?

2.3 How many visitors make a purchase?

- Out of 1979 potential customers who visit the site, 361 became customers.
- Although 1881 visitors added products in their shopping cart, apparently 1520 forgot about them and left without buying anything.

```
1   -- 2.3 Count visitors who made a purchase
2   SELECT page_name AS 'Page Name',
3          COUNT(DISTINCT user_id) AS 'Visitors'
4   FROM page_visits
5   GROUP BY page_name;
```

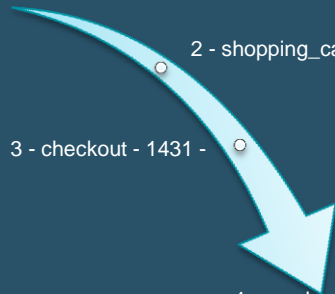| Page Name	Visitors
1 - landing_page	1979
2 - shopping_cart	1881
3 - checkout	1431
4 - purchase	361

VISITORS' JOURNEY

1- landing_page - 1979 -

2 - shopping_cart - 1881 -

3 - checkout - 1431 -

4 - purchase - 361 -

2. What is the user journey?

2.4 How many last touches on the purchase page is each campaign responsible for?

○ Worth noting that out of 622 potential customers who discovered the site through Medium (*interview-with-cool-tshirts-founder*), only 7 of them made a final purchase.

```sql
-- 2.4 Count last touches per campaign and source that led to a purchase

/* Create a temporary table that returns the last touches for each user_id */
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),
/* Create another temporary table adding source and campaign columns by
joining them on user_id and timestamp */
    lt_attr AS (
    SELECT lt.user_id,
        lt.last_touch_at,
        pv.utm_source,
        pv.utm_campaign
    FROM last_touch AS lt
    JOIN page_visits AS pv
        ON lt.user_id = pv.user_id
        AND lt.last_touch_at = pv.timestamp)
/* Create a table that includes the total number of last touches group by
source and campaign */
SELECT lt_attr.utm_source AS 'Source',
    lt_attr.utm_campaign AS 'Campaign',
    COUNT(*) AS 'Count'
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

Source	Campaign	Purchases
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. What is the user journey?

2.4 How many last touches on the purchase page is each campaign responsible for?

- 46.82% of purchases were generated via email newsletter and email retargeting.

```sql
1   -- 2.4 Count last touches per campaign and source that led to a purchase (in percentage terms)
2
3   /* Create a temporary table that returns the last touches for each user_id */
4   WITH last_touch AS (
5       SELECT user_id,
6           MAX(timestamp) AS last_touch_at
7       FROM page_visits
8       WHERE page_name = '4 - purchase'
9       GROUP BY user_id),
10  /* Create another temporary table adding source and campaign columns by joining them on user_id and timestamp */
11      lt_attr AS (
12      SELECT lt.user_id,
13          lt.last_touch_at,
14          pv.utm_source,
15          pv.utm_campaign
16      FROM last_touch AS lt
17      JOIN page_visits AS pv
18        ON lt.user_id = pv.user_id
19       AND lt.last_touch_at = pv.timestamp),
20  /* Create a temporary table that includes the total number of last touches group by source and campaign */
21      counting AS (
22      SELECT lt_attr.utm_source AS source,
23          lt_attr.utm_campaign AS campaign,
24          COUNT(*) AS lt_count
25      FROM lt_attr
26      GROUP BY 1, 2
27      ORDER BY 3 DESC),
28  /* Create a temporary table that includes the total number of last touches */
29      total_counting AS (
30      SELECT SUM(counting.lt_count) AS lt_total
31      FROM counting),
32  /* Create a temporary table that is a CROSS JOIN between counting table and total_counting table */
33      cj AS (
34      SELECT *
35      FROM counting
36      CROSS JOIN total_counting)
37  /* Create a table that includes the total number of last touches in percentage terms group by source and campaign */
38  SELECT cj.source AS 'Source',
39      cj.campaign AS 'Campaign',
40      ROUND(100.0 * cj.lt_count / cj.lt_total,2) AS '% Purchases'
41  FROM cj
42  GROUP BY 1, 2
43  ORDER BY 3 DESC;
```

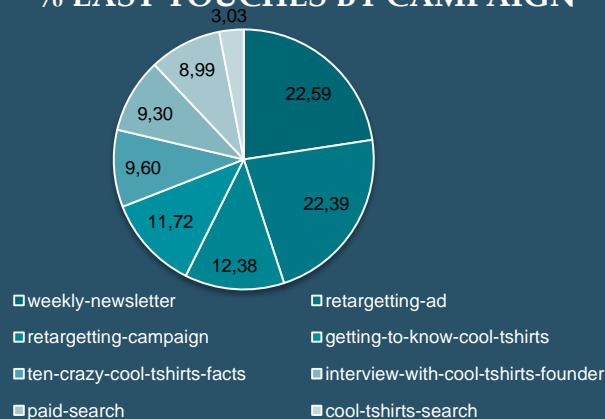| Source	Campaign	% Purchases
email	weekly-newsletter	31.86
facebook	retargetting-ad	31.3
email	retargetting-campaign	14.96
google	paid-search	14.4
buzzfeed	ten-crazy-cool-tshirts-facts	2.49
nytimes	getting-to-know-cool-tshirts	2.49
medium	interview-with-cool-tshirts-founder	1.94
google	cool-tshirts-search	0.55

2. What is the user journey?

2.5 What is the typical user journey?

o In the beginning of this journey, *CoolTShirts* is creating a connection with his potential customers through storytelling campaigns, such as *interview-with-cool-tshirts-founder* or *getting-to-know-cool-tshirts.*

o As we can see, these story campaigns attract 91,43% of potential customers and only 8,54% are driven to site through an organic search (in this case Google is our search engine).

% LAST TOUCHES BY CAMPAIGN

3,03
8,99
9,30
9,60
11,72
12,38
22,59
22,39

☐ weekly-newsletter   ☐ retargetting-ad
☐ retargetting-campaign   ☐ getting-to-know-cool-tshirts
☐ ten-crazy-cool-tshirts-facts ☐ interview-with-cool-tshirts-founder
☐ paid-search   ☐ cool-tshirts-search

% FIRST TOUCHES BY CAMPAIGN

8,54
29,11
31,43
30,92

☐ cool-tshirts-search   ☐ interview-with-cool-tshirts-founder
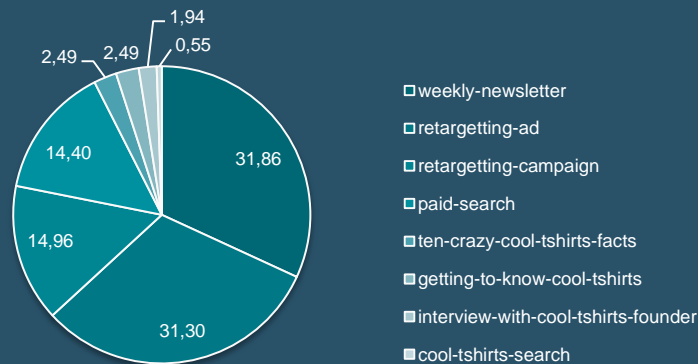☐ getting-to-know-cool-tshirts ☐ ten-crazy-cool-tshirts-facts
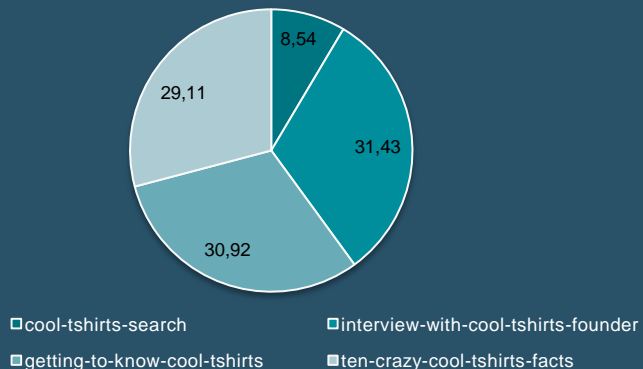
o Out of the 1979 visitors who discovered the site through storytelling campaigns, only 30.62% of them returned.

o In order to convert the visitors into purchasing customers is necessarily to remind them why they were interested in the first place.

12

2. What is the user journey?

o As we can see, only 6.92% of the visitors convert on their first visit to CoolTshirts.

o The 3 retargeting campaigns (*weekly-newsletter*, *retargeting-campaign* and *retargeting-ad*) attracted the right customers and generated most of the revenue.

o 78.12% of purchases were generated via email and Facebook retargeting ads.

% PURCHASES BY CAMPAIGN

- weekly-newsletter
- retargetting-ad
- retargetting-campaign
- paid-search
- ten-crazy-cool-tshirts-facts
- getting-to-know-cool-tshirts
- interview-with-cool-tshirts-founder
- cool-tshirts-search

1,94
0,55
2,49
2,49
14,40
31,86
14,96
31,30

3. Optimize the campaign budget

If CoolTShirts can re-invest in only 5 campaigns out of 8, I think it would be best to choose the following:
- the 3 storytelling campaigns that drove the visitors to *CoolTShirts* for the first time; even though these campaigns didn't generate revenue, they know how to captivate their audience and they should be continued;
- and 2 of the retargeting campaigns, *weekly-newsletter* and *retargetting-ad* because they are exceptional at winning back the visitors and generated most of the revenue (63.16%).

Source	Campaigns	1st Touches	% 1st Touches	Last Touches	% Last Touches	Purchases	% Purchases
mediu	interview-with-cool-tshirts-founder	622	31,43	184	9,30	7	1,94
buzzfeed	getting-to-know-cool-tshirts	612	30,92	232	11,72	9	2,49
nytimes	ten-crazy-cool-tshirts-facts	576	29,11	190	9,60	9	2,49
google	cool-tshirts-search	169	8,54	60	3,03	2	0,55
email	weekly-newsletter	0	0	447	22,59	115	31,86
facebook	retargetting-ad	0	0	443	22,39	113	31,30
email	retargetting-campaign	0	0	245	12,38	54	14,96
google	paid-search	0	0	178	8,99	52	14,40