



CoolTShirts

Learn SQL from Scratch

Ana Vieira Bouça

2018.08.07

Table of Contents

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?

Introduction I

Web developers, marketers, and data analysts can track how users arrive at their site using UTM parameters. If an ad campaign drives a lot of visits to their site, then they know that campaign is working! We say that those visits are attributed to the ad campaign.

CoolTShirts sells shirts of all kinds, as long as they are T-shaped and cool. Recently, CTS started a few marketing campaigns to increase website visits and purchases.

Using touch attribution, they'd like to map their customers' journey: from initial visit to purchase.

page_visits	
A table describing each time a user visits the CoolTShirts website	
Column	Description
user_id	A unique identifier for each visitor to a page
timestamp	The time at which the visitor came to the page
page_name	The title of the section of the page that was visited
utm_source	Identifies which site sent the traffic (i.e., google, newsletter, or facebook_ad)
utm_medium	Identifies what type of link was used (i.e., cost-per-click or email)
utm_campaign	Identifies the specific ad or email blast (i.e., june-21-newsletter or memorial-day-sale)

Introduction II

- A **UTM** code is a simple code that you can attach to a custom URL to track a source, media, and campaign name. A common use of the UTM code is to create a custom URL for each offline campaign and then redirect that URL to any forwarding address that you assign to it, probably your primary domain. This will allow you to track how a specific source is working without the need to create custom landing pages for each campaign;
- This way, when we consider first touch, it is just a first source for every customer. This is a good way to know how visitors are found on a website, because it considers only the first utm_source of each client, like facebook or buzzfeed, in this particular case;
- On the other hand, last touch assignment is only one last source for each client. This is a good way to know how visitors are attracted to a site, especially to make a final purchase, as it considers only the last utm_source of each client, like e-mail;
- Besides **utm_source**, we also use **utm_campaign**, since it allows to identify the specific ad or e-mail blast.



1. Get familiar with CoolTShirts

How many campaigns and sources does CoolTShirts use? Which source is used for each campaign?

- As we can see, we have 8 different campaigns, as specific messages to particular groups, with many specific sources used, as the type of communication in a campaign;
- We also have the same source running across two different campaigns, like paid-search or cool-tshirts-search throughout google source;
- Professional marketers generally use different kinds of sources, trying to get to specific groups. As it shows, we have a source like facebook, but also a classic one, like nytimes.

```
1  --1. Count the number of campaigns, sources and related them:
2
3  -- Number of distinct campaigns
4
5  SELECT COUNT(DISTINCT utm_campaign) as 'Campaign Count'
6  FROM page_visits;
7
8  -- Number of distinct sources
9
10 SELECT COUNT(DISTINCT utm_source) as 'Source Count'
11 FROM page_visits;
12
13 -- Relationship between a campaign and its sources
14
15 SELECT DISTINCT utm_campaign as 'Campaign',
16 utm_source as 'Source'
17 FROM page_visits;
18
```

Campaign Count

8

Source Count

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. Get familiar with CoolTShirts

What pages are on the CoolTShirts website?

- A function using the **distinct** function returns unique values, filtering the visitors in duplicate;
- This way, we can see the actual names of the 4 identified pages ('landing_page', shopping_cart', 'checkout', and 'purchase'), as the subset of actual visited pages on the website;
- Through these different type of pages, we can verify the visitor's unique journey, as ahead, since we can identify, for example, the type of **source** which the visitor used to get there, and try to make a pattern, that will help to explore and improve the campaigns, adapting them in the best way.

```
20 --2. Distinct pages in the ColltShirts website
21
22 SELECT DISTINCT page_name as 'Page Name'
23 FROM page_visits;
```

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

2. What is the user journey?

How many first touches is each campaign responsible for?

- Through the **minimum timestamps** related to each of the visitors, it is possible to determine the source and the responsive campaign that stimulated the first contact with the website;
- It was necessary to create a temporary table of **first_touch attribution**, in order to consider only the first source for each visitor/customer;
- In addition, of the 8 campaigns initially identified in the dataset, only 4 are determinant for the first contact. This will also be analyzed later.

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

```
27 --3. Count the number of first touches per campaign
28
29 --Temporary table with first touches by user id
30
31 WITH first_touch AS (SELECT user_id, MIN(timestamp) as first_touch_at
32                       FROM page_visits
33                       GROUP BY user_id),
34
35 --Temporary table with timestamp and user_id from page_visits
36
37 ft_attr AS (
38     SELECT ft.user_id,
39           ft.first_touch_at,
40           pv.utm_source,
41           pv.utm_campaign
42     FROM first_touch ft
43     JOIN page_visits pv
44       ON ft.user_id = pv.user_id
45       AND ft.first_touch_at = pv.timestamp)
46
47 --Count the number of rows where a specific source is associated with the
   same campaign
48
49 SELECT ft_attr.utm_source AS 'Source',
50       ft_attr.utm_campaign AS 'Campaign',
51       COUNT(*) AS 'Count'
52 FROM ft_attr
53 GROUP BY 1, 2
54 ORDER BY 3 DESC;
```

2. What is the user journey?

How many last touches is each campaign responsible for?

- As before, a temporary table was created, but this time, the **last_touch attribution**;
- In that case, the **maximum timestamps** related to each of the visitors were identified, determining how they were drawn back to the website;
- Also, it is possible to identify that the campaigns that run the best results in first touch attribution, were the least significant in last touch attribution. And they have different types of sources.

Source	Campaign	Count
email	weekly-newsletter	447
facebook	retargeting-ad	443
email	retargeting-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

```
56 --4. Count the number of last touches per campaign
57
58 --Temporary table with last touches by user id
59
60 WITH last_touch AS (SELECT user_id, MAX(timestamp) as last_touch_at
61                      FROM page_visits
62                      GROUP BY user_id),
63
64 --Temporary table with timestamp and user_id from page_visits
65
66 lt_attr AS (
67   SELECT lt.user_id,
68          lt.last_touch_at,
69          pv.utm_source,
70          pv.utm_campaign
71   FROM last_touch lt
72   JOIN page_visits pv
73     ON lt.user_id = pv.user_id
74     AND lt.last_touch_at = pv.timestamp)
75
76 --Count the number of rows where a specific source is associated with the
77 same campaign
78
79 SELECT lt_attr.utm_source AS 'Source',
80        lt_attr.utm_campaign AS 'Campaign',
81        COUNT(*) AS 'Count'
82 FROM lt_attr
83 GROUP BY 1, 2
84 ORDER BY 3 DESC;
```


2. What is the user journey?

How many visitors make a purchase?

- Of the 1979 visitors, only 361 made a purchase, which represents only 18% of 'becoming a customer';
- Quite simply, a reassessment of the campaigns, as well as the associated budget, will have to be made.

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

- Identifying in the dataset, the maximum timestamp, where the page visited was a '4 - purchase', we can identify the sources and campaigns responsible for the conversion;
- Once more, they are not the same utilized in first touch campaigns;
- Apparently, some changes must be made in the company.

Source	Campaign	Count
email	weekly-newsletter	115
facebook	retargeting-ad	113
email	retargeting-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

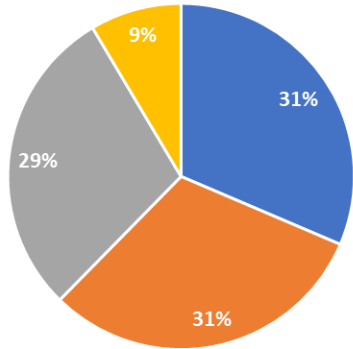
Visitors that made a purchase

361

```
85 --5.Count the number of visitors who made purchases
86
87 SELECT COUNT(DISTINCT user_id) AS 'Visitors that made a purchase'
88 FROM page_visits
89 WHERE page_name = '4 - purchase';
90
91 --6. Count how many last touches on the purchase page is each campaign
92 --responsible for
93
94 --Temporary table with last touches by user id in page '4 - purchase'
95
96 WITH last_touch AS (
97     SELECT user_id,
98           MAX(timestamp) AS last_touch_at
99     FROM page_visits
100    WHERE page_name = '4 - purchase'
101    GROUP BY user_id),
102
103 --Temporary table with timestamp and user_id from page_visits
104
105 lt_attr AS (
106     SELECT lt.user_id,
107           lt.last_touch_at,
108           pv.utm_source,
109           pv.utm_campaign
110     FROM last_touch lt
111    JOIN page_visits pv
112      ON lt.user_id = pv.user_id
113     AND lt.last_touch_at = pv.timestamp)
114
115 --Count the number of rows where a specific source is associated with
116 --the same campaign
117
118 SELECT lt_attr.utm_source AS 'Source',
119       lt_attr.utm_campaign AS 'Campaign',
120       COUNT(*) AS 'Count'
121 FROM lt_attr
122 GROUP BY 1, 2
123 ORDER BY 3 DESC;
```

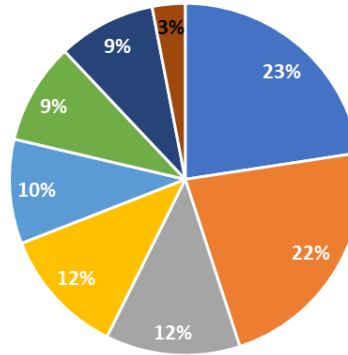
2. What is the user journey?

% FIRST TOUCHES BY CAMPAIGN



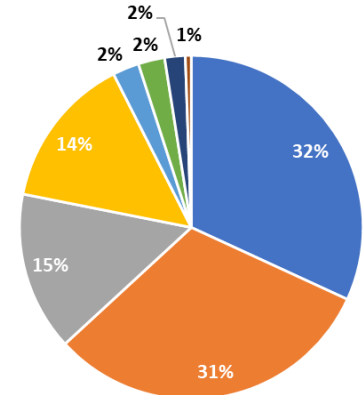
- interview-with-cool-tshirts-founder
- getting-to-know-cool-tshirts
- ten-crazy-cool-tshirts-facts
- cool-tshirts-search

% LAST TOUCHES BY CAMPAIGN



- 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

% OF VISITORS THAT MADE A PURCHASE



- 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

2. What is the user journey?

What is the typical user journey?

- In this dataset, we have 2 types of campaigns. First, **Storytelling Campaigns**, which can go through all forms of marketing and social networking is where this mode of storytelling based on experiences and short stories has grown, like 'getting-to-know-cool-tshirts' or 'ten-crazy-cool-tshirts-facts';
- However, we also, have **Redirect** (or remarketing) **Campaigns**, which are a form of online advertising that can help you be flagged in front of returned traffic after your visitors have left the site, as 'weekly-newsletter' or 'retargeting-ad';
- **'%First Touches by Campaign'** → of the 4 campaigns, 3 are Storytelling type and represent 91% of the first touches;
- **'% Last Touches by Campaign'** → there's evidence of last touches in all the campaigns in the dataset. However, and despite of having 57% of remarketing campaigns, which sounds adequate for this instance, there's still 31% of storytelling campaigns, which as shown before, are the least responsible for last touches;
- **'% of Visitors that made a purchase'** → Once again, data showed that remarketing campaigns are the most significant when a visitor becomes a customer!! As it represents 78%. Still, the company invest 6% in storytelling.

3. Optimize the Campaign budget

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

- If the choice actually went through 5 of the 8 campaigns, I would select the 3 most incident in the first touch attribution - the green ones - , and the 2 more incident in purchase conversion - the blue ones;
 - However, as mentioned above, it is noted that **retargeting campaigns** are fundamental in the conversion to purchase, so *retargeting-campaign* should also be considered;
- The other two campaigns - in red - can be restructured, especially in the sense of increased conversion from visitor to customer, since the proportion is very low (18%);

CAMPAIGN	FIRST TOUCHES	% FIRST TOUCHES	LAST TOUCHES	% LAST TOUCHES	PURCHASES	% PURCHASES
interview-with-cool-tshirts-founder	622	31%	184	9%	7	2%
getting-to-know-cool-tshirts	612	31%	232	12%	9	2%
ten-crazy-cool-tshirts-facts	576	29%	190	10%	9	2%
cool-tshirts-search	169	9%	60	3%	2	1%
weekly-newsletter	0	0%	447	23%	115	32%
retargeting-ad	0	0%	443	22%	113	31%
retargeting-campaign	0	0%	245	12%	54	15%
paid-search	0	0%	178	9%	52	14%

3. Optimize the Campaign budget

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

- In addition, in relation to orange campaigns, the budget for **storytelling campaigns** for conversion into purchase should be redirected to the rest, reinforcing, for example, *retargeting-campaign* in the last touch attribution, thus increasing the likelihood of conversion in purchase;
- Finally, other considerations can be studied in order to improve the performance of the campaigns, namely the type of sources used, the characteristics of each campaign and the selection of the appropriate source, as well as the consent of utm_medium, identifying the type of link used.

CAMPAIGN	FIRST TOUCHES	% FIRST TOUCHES	LAST TOUCHES	% LAST TOUCHES	PURCHASES	% PURCHASES
interview-with-cool-tshirts-founder	622	31%	184	9%	7	2%
getting-to-know-cool-tshirts	612	31%	232	12%	9	2%
ten-crazy-cool-tshirts-facts	576	29%	190	10%	9	2%
cool-tshirts-search	169	9%	60	3%	2	1%
weekly-newsletter	0	0%	447	23%	115	32%
retargeting-ad	0	0%	443	22%	113	31%
retargeting-campaign	0	0%	245	12%	54	15%
paid-search	0	0%	178	9%	52	14%



CoolTShirts

Learn SQL from Scratch

Ana Vieira Bouça

2018.08.07

End.