Capstone

CoolTShirts - Attribution Theories

Learn SQL from Scratch
By Mahsa Dinyari
April 1, 2019 - May 27, 2019

Table of Contents

I. Get familiar with CoolTShirts
- ➔ *How many campaigns and sources does CoolTShirts use, and how are they related?*
- ➔ *The difference between utm_campaign and utm_source*
- ➔ *What pages are on the CoolTShirts website?*

II. What is the user journey throughout CoolTshirts' website?
- ➔ *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?*

III. Optimize the Campaign
- ➔ *An explanation of the 5 campaigns CoolTShirts should re-invest in, based on user data found in the previous sections!*

I. Get Familiar with CoolTShirts

Results

COUNT(DISTINCT utm_campaign)	
8	

COUNT(DISTINCT utm_source)	
6	

utm_campaign	utm_source
getting-to-know-cool-ts hirts	nytimes
weekly-newsletter	email
Ten-crazy-cool-tshirts-facts	buzzfeed
retargetting-campaign	email
retargetting-ad	facebook
interview-with-cool-tsh irts-founder	medium
paid-search	google
cool-tshirts-search	google

We use **DISTINCT** to find count the number of DIFFERENT (not repeated) campaigns and sources.

Query

```
1   SELECT COUNT(DISTINCT utm_campaign)
2   FROM page_visits;
3
4   SELECT COUNT(DISTINCT utm_source)
5   FROM page_visits;
6
7   SELECT DISTINCT utm_campaign, utm_source
8   FROM page_visits;
```

How many campaigns does CoolTShirt use? Which source is used for each campaign?

➔ The source (recognized as **utm_source** within the user data) is the platform or vendor where traffic originates. For example, Twitter, Facebook, Google, Pinterest.

➔ The campaign (**utm_campaign**) is the specific ad or email blast that drew a visitor to a website. For example, a company's *November Newsletter*, or a specific Retargetting Ad.

➔ A source can run many different campaigns. And a specific campaign can also be run across various sources. For example, both Facebook and Pinterest can run the same exact Retargetting Ad.

➔ Buzzfeed (the **source**) ran the article "Ten Crazy CoolTShirts Facts" (the **campaign**) on its website. This is how 582 different (**distinct**) users got to CoolTShirts' landing page

How many pages are on the CoolTshirts website?

➔ We are **NOT** looking for a long list of page names based on every individual visitor's experience - there would be <u>many</u> repeats.

➔ We **ARE** looking to know what the different (**distinct**) page names are, to make analysis simple.

```
1    SELECT DISTINCT page_name AS 'Page Name'
2    FROM page_visits;
```
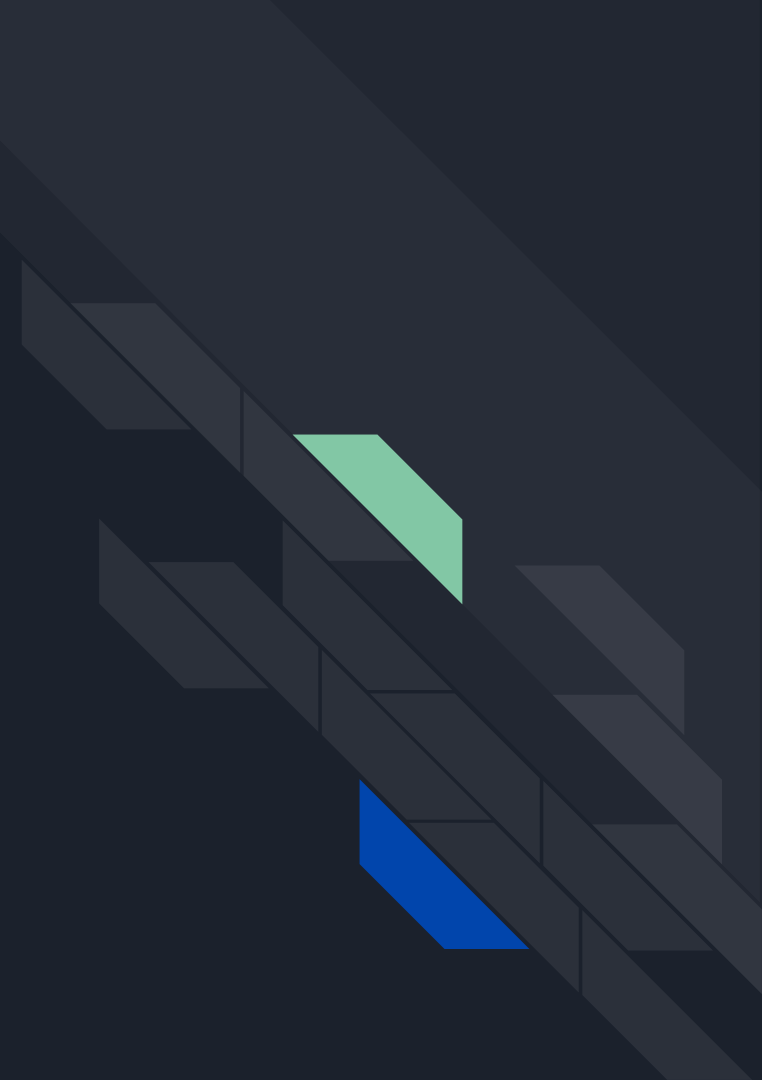
Query

There are 4 different page names within CoolTshirts' user data. If we didn't make the query **distinct**, these would be repeated several times and difficult to identify.

I renamed page_name using **AS** to make the title easier to read.

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

Results

II. What is the user journey throughout CoolTShirts' website?

How many first touches is each campaign responsible for?

We are first creating a temporary table called **first_touch**

```
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    SELECT ft.user_id,
7        ft.first_touch_at,
8        pv.utm_source,
9        pv.utm_campaign,
10       COUNT(utm_campaign)
11   FROM first_touch ft
12   JOIN page_visits pv
13       ON ft.user_id = pv.user_id
14       AND ft.first_touch_at = pv.timestamp
15   GROUP BY utm_campaign
16   ORDER BY 5 DESC;
```

To identify the first click that initially brought a user to the CoolTShirts website, we use the **MIN(timestamp)** from the already existing **page_visits** table. We call this value "first_touch_at" in the new **first_touch** table that we just created.

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-cooltshirts-facts	576
99684	2018-01-13 13:20:49	google	cool-tshirts-search	169

Here, we can see which **campaign** best served the **first touch attribution** of users leading them to the CoolTShirts website : the *Interview with CoolTShirts' founder.*

This information is crucial for CoolTshirts' development because it identifies which **campaign** drew the most users to the website. The company may want to use this *Interview* across more **sources** (described in Section III).

How many last touches is each campaign responsible for?

```sql
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;
```

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	retargetting-ad	443
99990	2018-01-16 11:35:09	email	retargetting-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

I ordered CoolTShirts' last touch attribution in descending (**DESC**) order.

This allows us to see that the *Weekly Newsletter* (the **campaign**) sent via email (the **source**) is how the most users (what the **COUNT** feature describes) are drawn back to the website.

How many visitors make a purchase?

```
1   SELECT COUNT(DISTINCT user_id) AS '# of visitors
2   on landing page'
3   FROM page_visits
4   WHERE page_name = '1 - landing_page';
```

```
1   SELECT COUNT(DISTINCT user_id) AS '# of visitors
    making purchase'
2   FROM page_visits
3   WHERE page_name = '4 - purchase';
```

# of visitors on landing page
1979

# of visitors making purchase
361

Compared to the number of **distinct** users who make it to CoolTShirts' landing page, there is a significant drop in users making an actual purchase.

This is an 81.76 % drop, meaning CoolTShirts needs to find ways to optimize their sales via their website (described further in Section III)

How many last touches on the *Purchase Page* is each campaign responsible for?

```
1   WITH last_touch AS (
2     SELECT user_id,
3     MAX(timestamp) as last_touch_at
4     FROM page_visits
5     WHERE page_name = '4 - purchase'
6     GROUP BY user_id)
7   SELECT lt.user_id,
8     lt.last_touch_at,
9     pv.utm_source,
10    pv.utm_campaign,
11    COUNT(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  GROUP BY utm_campaign
17  ORDER BY 5 DESC;
```

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	retargetting-ad	113
99285	2018-01-24 09:00:58	email	retargetting-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

➔ This specifies the **last touches** of users to the *Purchase Page*

➔ These results (grouped by campaign) identify the **sources** and **campaigns** that warranted the most users to purchase from CoolTshirts

➔ The closer users get to the *Purchase Page*, the more users drop off. The click through rate continues to diminish until the final page is reached.
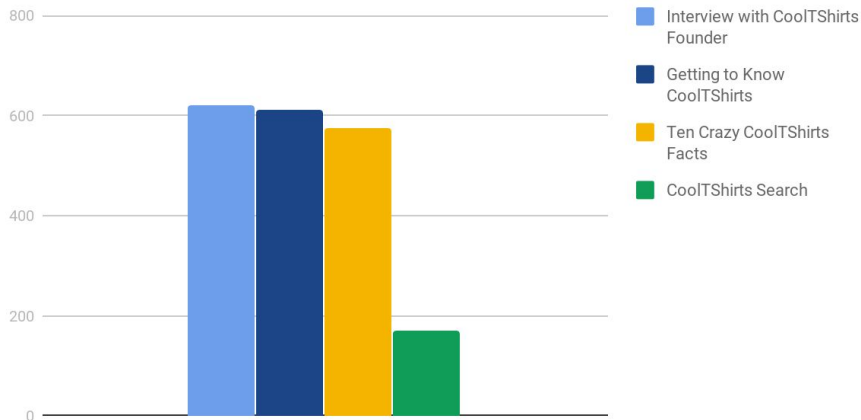
What is the typical user journey?

It's worth noting that 3 campaigns significantly surpass the other in terms of **first touch attribution**.

91% of users are attracted to the website through campaigns that allow them to *get to know the company* and who runs it. Just searching for CoolTShirts in Google wasn't nearly as effective.
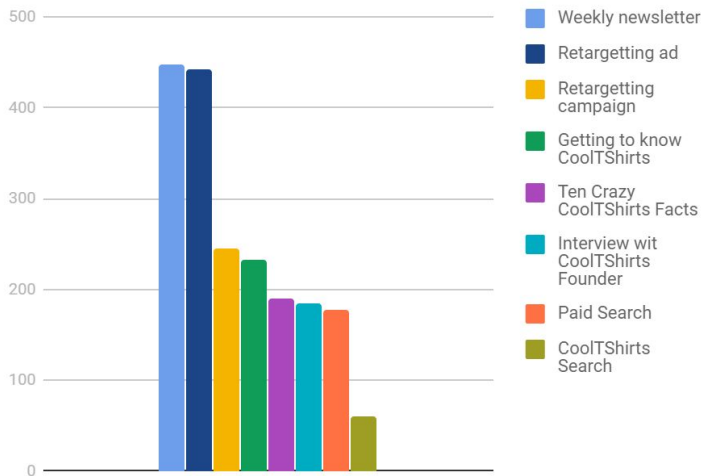
First Touch (By Campaign)
By # of users

Legend:
- Interview with CoolTShirts Founder
- Getting to Know CoolTShirts
- Ten Crazy CoolTShirts Facts
- CoolTShirts Search

Last Touch by Campaign
By # of users

Legend:
- Weekly newsletter
- Retargetting ad
- Retargetting campaign
- Getting to know CoolTShirts
- Ten Crazy CoolTShirts Facts
- Interview wit CoolTShirts Founder
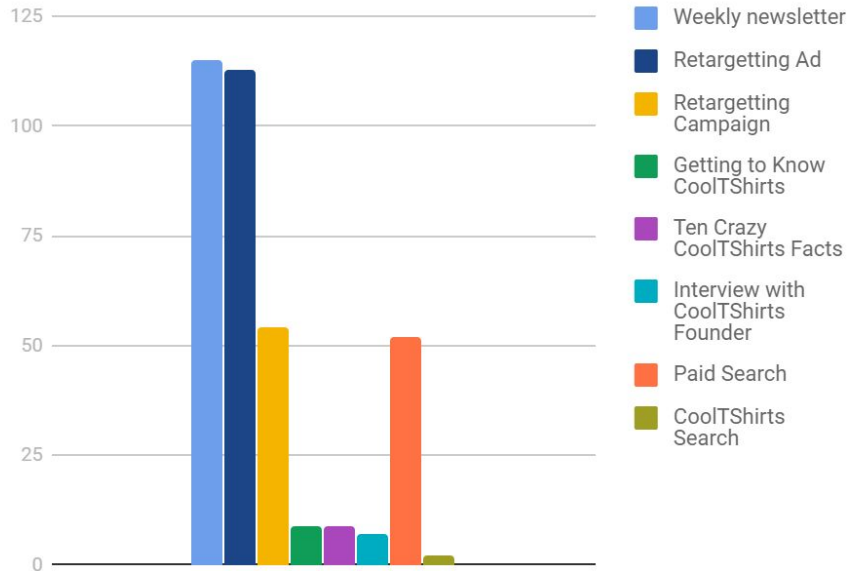- Paid Search
- CoolTShirts Search

The campaigns that initially drew visitors to the website do not attract nearly as many visitors back to the site.

45% of visitors returned after exposure to the *Weekly Newsletter* (sent via email) or the Retargetting Ad (on Facebook).

It is clear that even the most effective campaigns are effective for different purposes.

Last Touch on Purchasing Page (by Campaign)

By # of users

Weekly newsletter
Retargetting Ad
Retargetting Campaign
Getting to Know CoolTShirts
Ten Crazy CoolTShirts Facts
Interview with CoolTShirts Founder
Paid Search
CoolTShirts Search

Not only are the **Weekly Newsletter** and the **Retargetting Ad** the two campaigns that attract the most users <u>back </u>to the website . . .

They also prompt most of the 361 users that return to the site (63% of them) to the final page : the *Purchasing Page*

Still, the campaigns that drew users back to the website are somewhat different than the campaigns that motivated users to actually make a purchase.

The big picture : the company needs to conduct further analysis to determine which campaigns are most influential overall.

What is the typical user journey?

Optimizing the Campaign Budget

What are the **5 campaigns** CoolTShirts should re-invest in?

Based on First - Touch Attribution

➔ 3 campaigns that initially attracted the most users (91%) to the CoolTShirts website should be re-invested in

➔ What these campaigns have in common is that they allow the customer to get to know the company through informative articles

➔ Statistically, it would be best to re-invest in the *Interview with CoolTShirts Founder* and *Getting to know CoolTShirts*

➔ Reinvesting in *Ten Crazy CoolTShirts Facts* would have similar upturn, as it captivates almost as many users as the other two articles

Based on Last - Touch Attribution AND Last - Touch on Purchase Page

➔ According to the number of Last - Touches that each campaign is responsible for alone, the *Weekly Newsletter* and the *Retargetting Ad* should be re-invested in

➔ According to the number of Last - Touches to the *Purchase Page*, the **Weekly Newsletter** and the **Retargetting Ad** are still the leading campaigns directing the most users to make a purchase

◆ 63% of all Last - Touches to the *Purchase Page* are attributed to one of these two campaigns. The other 37% are dispersed among the 6 other campaigns, making the *Newsletter* and the *Ad* the clear leaders

◆ Since attracting users to actually make a purchase is how CoolTShirts will ultimately become profitable, these two campaigns are just as important as the campaigns that direct First - Touches to the site