



# CAPSTONE PROJECT - ATTRIBUTION

Learn SQL from Scratch

Alexandra Massam

August 8, 2018

# Table of Contents

1. Get Familiar with the Company
2. What is the user journey?
3. Optimize the campaign budget

# 1. Get Familiar With the Company

# 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`.

There are 8 campaigns and 6 sources used. Each source is the website in which each campaign is located, and the campaign is the point at which each user clicked off of the site to access the CoolTShirts site. The source identifies which site sent the traffic, and the campaign identifies the specific ad or e-mail blast. See the table below for the data used to support my claim.

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

```
SELECT DISTINCT(utm_source)
FROM page_visits;
```

```
SELECT DISTINCT(utm_campaign)
FROM page_visits;
```

```
SELECT DISTINCT utm_source, utm_campaign
FROM page_visits;
```

# Get Familiar With the Company

## What pages are on their website?

The pages on their website are the landing page, shopping cart, checkout, and purchase pages. See the table below for the data used to support my claim.

page_name
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

```
SELECT DISTINCT page_name
FROM page_visits;
```

## 2. What is the user journey?

# What is the user journey?

## How many first touches is each campaign responsible for?

Here are the campaigns and the number of first touches each is responsible for: interview-with-cool-tshirts-founder is responsible for 622, getting-to-know-cool-tshirts for 612, ten-crazy-cool-tshirts-facts for 576 and cool-tshirts-search for 169. See table below for the data used to support my claim.

ft.attr.utm_source	ft.attr.utm_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

```
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,
       ft_attr.utm_campaign,
       COUNT(*)
FROM ft_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

# What is the user journey?

## How many last touches is each campaign responsible for?

The campaigns and last touches are as follows: weekly-newsletter has 447, retargeting-ad has 443, retargeting-campaign has 245, getting-to-know-cool-tshirts has 232, ten-crazy-cool-tshirts-facts has 190, interview-with-cool-tshirts-founder has 184, paid-search has 178, and cool-tshirts-search has 60. See table below for the data used to support my claim.

lt.attr.utm_source	lt.attr.utm_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

```
WITH last_touch AS (SELECT user_id,
    MAX(timestamp) as last_touch_at
    FROM page_visits
    GROUP BY user_id),
lt_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 lt_attr.utm_source,
    lt_attr.utm_campaign,
    COUNT(*)
FROM lt_attr
GROUP BY utm_campaign
ORDER BY 3 DESC;
```



# What is the user journey?

**How many visitors make a purchase?**

361 visitors made purchases.

COUNT(*)
361

```
SELECT COUNT(*)  
FROM page_visits  
WHERE page_name = '4 - purchase';
```

# What is the user journey?

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

Last touches on the purchase page are as follows: weekly-newsletter has 115, retargeting-ad has 113, retargeting-campaign has 54, paid-search has 52, getting-to-know-cool-tshirts has 9, ten-crazy-cool-tshirts-facts also has 9, interview-with-cool-tshirts-founder has 7, and cool-tshirts-search has 2. See the table below for the data used to support my claim.

lt.attr.utm_source	lt.attr.utm_campaign	COUNT(*)
email	weekly-newsletter	115
facebook	retargeting-ad	113
email	retargeting-campaign	54
google	paid-search	52
nytimes	getting-to-know-cool-tshirts	9
buzzfeed	ten-crazy-cool-tshirts-facts	9
medium	interview-with-cool-tshirts-founder	7
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),
lt_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 lt_attr.utm_source,
    lt_attr.utm_campaign,
    COUNT(*)
FROM lt_attr
GROUP BY utm_campaign
ORDER BY 3 DESC;
```

# What is the user journey?

## What is the typical user journey?

The typical use journey is to begin at the source, a webpage like the Buzzfeed, then to click on the campaign, like a retargeting-ad or weekly-newsletter, which takes them to the CoolTShirts landing page. They then pick what they would like to purchase and place it in the shopping cart, they are taken to the checkout, and finally they make the purchase on the purchase page. See the table below for the data used to support my claim. The table is only a sample of the entirety of the data used, and is meant to show a pattern.

```
SELECT *  
FROM page_visits  
ORDER BY user_id;
```

page_name	timestamp	user_id	utm_campaign	utm_source
1 - landing_page	2018-01-24 03:12:16	10006	getting-to-know-cool-tshirts	nytimes
2 - shopping_cart	2018-01-24 04:04:16	10006	getting-to-know-cool-tshirts	nytimes
3 - checkout	2018-01-25 23:10:16	10006	weekly-newsletter	email
1 - landing_page	2018-01-25 20:32:02	10030	ten-crazy-cool-tshirts-facts	buzzfeed
2 - shopping_cart	2018-01-25 23:05:02	10030	ten-crazy-cool-tshirts-facts	buzzfeed
3 - checkout	2018-01-28 13:26:02	10030	retargetting-campaign	email
4 - purchase	2018-01-28 13:38:02	10030	retargetting-campaign	email

### 3. Optimize the campaign budget

# Optimize the campaign budget

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

CoolTShirts should re-invest in the following 5 campaigns: weekly-newsletter, retargeting-ad, retargeting-campaign, paid-search and either getting-to-know-cool-tshirts or ten-crazy-tshirts-facts, which are tied for fifth place at 9 each. When I used the code to the right, the table displayed, in order from greatest to least, the number of times the campaigns that users clicked on that also ended in a purchase. That is why the aforementioned campaigns are the ones that should be re-invested in.

```
SELECT COUNT(utm_campaign), utm_campaign, user_id, page_name
FROM page_visits
WHERE page_name = '4 - purchase'
GROUP BY utm_campaign
ORDER BY 1 DESC;
```

COUNT(utm_campaign)	utm_campaign	page_name
115	weekly-newsletter	4 - purchase
113	retargeting-ad	4 - purchase
54	retargeting-campaign	4 - purchase
52	paid-search	4 - purchase
9	getting-to-know-cool-tshirts	4 – purchase
9	ten-crazy-cool-tshirts-facts	4 – purchase