



# Capstone: Attribution

Learn SQL from Scratch

Julie Floyd

July 22, 2018

# Table of Contents

1. Get familiar with CoolTShirts
2. What is the user journey?
3. Optimize the campaign budget

# 1. Get Familiar with CoolTShirts

# 1.1 page\_visits database table

Below is the page\_visits database table.

- page\_name is the title of the section of the page that was visited
- timestamp is the time that the visitor came to the page
- user\_id is the unique identifier for each visitor to the page
- utm\_campaign identifies the ad or email that included a link to the CoolTShirts website
- utm\_source identifies the source of the traffic (i.e., the site where the ad was placed and the link was clicked)

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
1 - landing_page	2018-01-05 18:31:17	10045	getting-to-know-cool-tshirts	nytimes
2 - shopping_cart	2018-01-05 21:16:17	10045	getting-to-know-cool-tshirts	nytimes
3 - checkout	2018-01-09 03:05:17	10045	retargetting-ad	facebook

# 1.1 Campaigns and sources used by CoolTShirts

How many campaigns and sources does CoolTShirts use?

- Number of distinct campaigns = 8
- Number of distinct sources = 6

The campaign is how the visitor got to the CoolTShirts website and the source is where the visitor was before coming to the CoolTShirts website. See the table below to know which source is used for each campaign:

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

```
/* Number of Distinct Campaigns Query */
SELECT COUNT (DISTINCT utm_campaign)
FROM page_visits;

/* Number of Distinct Sources Query */
SELECT COUNT (DISTINCT utm_source)
FROM page_visits;

/* How are the Campaigns and Sources Related Query */
SELECT DISTINCT utm_campaign,
                utm_source
FROM page_visits;
```

## 1.2 Pages on the CoolTShirts website

There are four different pages on the CoolTShirts website.

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

```
/* Pages on the CoolTShirts website Query */  
SELECT DISTINCT page_name  
FROM page_visits;
```

## 2. What is the User Journey?

## 2.3 First Touches

How many first touches is each campaign responsible for?

- See table below:

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

```
/* First Touch - Campaign Query */
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;
```



## 2.4 Last Touches

How many last touches is each campaign responsible for?

- See table below:

lt_attr.utm_source	lt_attr.utm_campaign	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

```
/* Last Touch - Campaign Query */
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 1, 2
ORDER BY 3 DESC;
```

## 2.5 Purchase

How many visitors make a purchase?

- 361

COUNT (DISTINCT user_id)
361

```
/* Purchase Query */  
SELECT COUNT (DISTINCT user_id)  
FROM page_visits  
WHERE page_name = '4 - purchase';
```

## 2.6 Last Touches = Purchase

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

- See table below:

lt_attr.utm_source	lt_attr.utm_campaign	COUNT(*)
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

```
/* Last Touch = Purchase - Campaign Query */
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 1, 2
ORDER BY 3 DESC;
```

## 2.7 Typical User Journey

- CoolTShirts receives the largest amount of visitors from medium with their “interview with CoolTShirts founder” campaign, with purchases resulting from only about 1% of those visits.
- CoolTShirts receives the second largest amount of visitors from the NY Times with their “getting to know CoolTShirts” campaign with purchases from about 2% of those visits.
- CoolTShirts has 1,979 unique user id’s in their customer database. For the weekly-newsletter email campaign that goes to all of their customers, about 23% of those customers make a purchase.
- The CoolTShirts conversation rate is about 18% - that means that out of the 1,979 first touches, 18% of them result in a customer purchase.

First touch source	First touch campaign	% purchase
medium	interview-with-cool-tshirts-founder	1%
nytimes	getting-to-know-cool-tshirts	2%
buzzfeed	ten-crazy-cool-tshirts-facts	2%
google	cool-tshirts-search	1%
Last touch source	Last touch campaign	% purchase
email	weekly-newsletter	23%
email	Retargeting-campaign	12%

- The weekly newsletter emails result in about twice as many purchases as the re-targeting campaign emails.
- Most of the purchases result from an email campaign to an existing customer.

## 2.7 Code for Additional Queries

- CoolTShirts has 1,979 unique user id's in their customer database. If they send an email campaign to all of the customers in their database, only about 23% of those result in a purchase.
- Only 5% of the site visitors leave before viewing their cart.
- 76% of the site visitors that view their shopping cart move on to the checkout process.
- Of that 76%, only 25% of them actually complete their purchase, which is not that bad!

```
/* Distinct User ID Query */
SELECT COUNT (DISTINCT user_id)
FROM page_visits;

/* User Journey by Campaign Query */
SELECT page_name,
       COUNT (page_name) as 'page visit count',
       utm_campaign,
       utm_source
FROM page_visits
GROUP BY utm_campaign,
         page_name;

/* User Journey by Page (Funnel) Query */
SELECT page_name,
       COUNT (DISTINCT user_id)
FROM page_visits
GROUP BY page_name;
```

# 3. Optimize the Campaign Budget

## 3.1 Optimize the Campaign Budget

CoolTShirts top Last Touch = Purchase sources are:

source	campaign	COUNT(*)
email	weekly-newsletter	115
facebook	retargeting-ad	113
email	retargeting-campaign	54

The First Touch source for the campaigns above is likely from these top First Touch sources:

source	campaign	COUNT(*)
medium	interview-with-cool-tshirts-founder	622
nytimes	getting-to-know-cool-tshirts	612
buzzfeed	ten-crazy-cool-tshirts-facts	576

## 3.1 Optimize the Campaign Budget

CoolTShirts needs to continue to invest in the sources that bring people to their site to browse, which is medium, nytimes, and buzzfeed. Once on the CoolTShirts site, it should be very easy for a visitor to sign up to receive CoolTShirts emails because that is the source of a large portion of their purchases. Therefore, the 5 campaigns in which CoolTShirts should re-invest are:

- **weekly-newsletter:** the majority of their purchases come from this campaign
- **retargeting-ad:** this receives the second highest number of purchases
- **retargeting-campaign:** this campaign receives the third highest number of purchases
- **interview-with-cool-tshirts-founder:** this is the largest source of visitor referrals, which can lead to signing up for emails or retargeting by Facebook ads and ultimately a purchase
- **getting-to-know-cool-tshirts:** this is the second largest source of visitor referrals, which can lead to signing up for emails or retargeting by Facebook ads and ultimately a purchase