



Attribution Queries

Capstone Project

Learn SQL from Scratch

Adam Rush

10/3/2018

Table of Contents

1. Getting Familiar with CoolTShirts
2. The User Journey
3. Optimizing the Campaign Budget

1. Getting Familiar with CoolTShirts

1.1 Getting Familiar with CoolTShirts

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

Q: How many distinct campaigns are there?

A: From the first query below, we find that the answer is **8** distinct campaigns.

* I took it a step further and wanted to see a list of the 8 distinct campaigns, so I wrote the second query below. The output is to the right.

```
/* How many distinct campaigns are there? */  
SELECT COUNT(DISTINCT utm_campaign)  
FROM page_visits;  
  
/* What are the distinct CoolTShirts campaigns? */  
SELECT DISTINCT utm_campaign AS 'campaign'  
FROM page_visits;
```

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

1.2 Getting Familiar with CoolTShirts

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

Q: How many distinct sources are there?

A: From the first query below, we find that the answer is **6** distinct sources.

* I took it a step further and wanted to see a list of the 6 distinct sources, so I wrote the second query below. The output is to the right.

```
/* How many distinct campaigns are there? */  
SELECT COUNT(DISTINCT utm_campaign)  
FROM page_visits;  
  
/* What are the distinct CoolTShirts traffic sources? */  
SELECT DISTINCT utm_source AS 'source'  
FROM page_visits;
```

source
nytimes
email
buzzfeed
facebook
medium
google

1.3 Getting Familiar with CoolTShirts

Q: How are the campaigns and sources related?

A: **utm_campaign** (the “campaign” column) identifies the specific ad or email blast in the campaign, whereas **utm_source** (the “source” column) identifies which site sent the traffic to the CoolTShirts website.

As you can see, you can have multiple campaigns coming from a single traffic source.

```
/* How are the campaigns and sources related? */  
SELECT DISTINCT utm_campaign AS 'campaign',  
               utm_source AS 'source'  
FROM page_visits;
```

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

1.4 Getting Familiar with CoolTShirts

Q: What pages are on the CoolTShirts website?

A: Using the query below, we are able to find a list of the 4 different pages on the CoolTShirts website. They are **Landing Page**, **Shopping Cart**, **Checkout** and **Purchase**.

```
/* What pages are on the CoolTShirts website? */  
SELECT DISTINCT page_name AS 'page'  
FROM page_visits;
```

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

2. The User Journey

2.1 The User Journey

Q: How many first touches is each campaign responsible for?

A: Using the query to the right, we are able to find a list of the 4 different campaigns that brought initial traffic to the CoolTShirts website, along with a count of the number of first touches for each campaign. We can see from the results that the campaign titled “interview-with-cool-tshirts-founder” brought in the most initial traffic to the site.

campaign	count
interview-with-cool-tshirts-founder	622
getting-to-know-cool-tshirts	612
ten-crazy-cool-tshirts-facts	576
cool-tshirts-search	169

```
/* How many first touches is each campaign responsible for? */
```

```
WITH first_touch AS (  
    SELECT user_id,  
           MIN(timestamp) AS first_touch_at  
    FROM page_visits  
    GROUP BY user_id)  
SELECT pv.utm_campaign AS 'campaign',  
       COUNT(*) AS 'count'  
FROM first_touch AS ft  
JOIN page_visits AS pv  
    ON ft.user_id = pv.user_id  
   AND ft.first_touch_at = pv.timestamp  
GROUP BY 1  
ORDER BY 2 DESC;
```

2.2 The User Journey

Q: How many last touches is each campaign responsible for?

A: Using the query below, we are able to find a list of the 8 different campaigns for CoolTShirts, along with a count of the number of last touches for each campaign. We can see from the results that the campaign titled “weekly-newsletter” was the highest in last touches for all campaigns.

```
/* How many last touches is each campaign responsible for? */
```

```
WITH last_touch AS (  
  SELECT user_id,  
         MAX(timestamp) AS last_touch_at  
  FROM page_visits  
  GROUP BY user_id)  
SELECT pv.utm_campaign AS 'campaign',  
       COUNT(*) AS 'count'  
FROM last_touch AS lt  
JOIN page_visits AS pv  
  ON lt.user_id = pv.user_id  
  AND lt.last_touch_at = pv.timestamp  
GROUP BY 1  
ORDER BY 2 DESC;
```

campaign	count
weekly-newsletter	447
retargeting-ad	443
retargeting-campaign	245
getting-to-know-cool-tshirts	232
ten-crazy-cool-tshirts-facts	190
interview-with-cool-tshirts-founder	184
paid-search	178
cool-tshirts-search	60

3. Optimizing the Campaign Budget

3.1 Optimizing the Campaign Budget

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

A: In order to answer this question fully, I decided to put together a table using SQL to find how many customers came through each page of the site, via each campaign.

My code is a little long, so it starts below, and continues in the column to the right. My results are on the next slide...

```
WITH lp AS(  
  SELECT utm_campaign AS 'campaign',  
         COUNT(*) AS 'landing_page'  
  FROM page_visits  
  WHERE page_name = '1 - landing_page'  
  GROUP BY 1),  
sc AS(  
  SELECT utm_campaign AS 'campaign',  
         COUNT(*) AS 'shopping_cart'  
  FROM page_visits  
  WHERE page_name = '2 - shopping_cart'  
  GROUP BY 1),
```

-- code continues to the right

-- continued from left

```
cp AS(  
  SELECT utm_campaign AS 'campaign',  
         COUNT(*) AS 'checkout'  
  FROM page_visits  
  WHERE page_name = '3 - checkout'  
  GROUP BY 1),  
pc AS(  
  SELECT utm_campaign AS 'campaign',  
         COUNT(*) AS 'purchase'  
  FROM page_visits  
  WHERE page_name = '4 - purchase'  
  GROUP BY 1)  
SELECT DISTINCT pv.utm_campaign AS 'campaign',  
               lp.landing_page AS 'landing_page',  
               sc.shopping_cart AS 'shopping_cart',  
               cp.checkout AS 'checkout',  
               pc.purchase AS 'purchase'  
FROM page_visits AS pv  
LEFT JOIN lp  
  ON lp.campaign = pv.utm_campaign  
LEFT JOIN sc  
  ON sc.campaign = pv.utm_campaign  
LEFT JOIN cp  
  ON cp.campaign = pv.utm_campaign  
LEFT JOIN pc  
  ON pc.campaign = pv.utm_campaign  
GROUP BY 1;
```

3.2 Optimizing the Campaign Budget

- Using this information, we can see that, while the “getting to know cool tshirts”, “interview with cool tshirts founder” and “ten crazy cool tshirts facts” articles brought in much of the traffic to the site, they were not effective at converting customers.
- On the other hand, the retargeting ad and campaign, as well as the weekly newsletter were great for getting customers to purchase, but not for bringing in any traffic to the site. What does this mean? Find out on the next slide!

campaign	landing_page	shopping_cart	checkout	purchase
cool-tshirts-search	171	133	7	2
getting-to-know-cool-tshirts	617	682	41	9
interview-with-cool-tshirts-founder	625	515	31	7
paid-search	-	-	179	52
retargeting-ad	-	-	445	113
retargeting-campaign	-	-	246	54
ten-crazy-cool-tshirts-facts	587	570	32	9
weekly-newsletter	-	-	450	115

3.3 Optimizing the Campaign Budget

- Looking at this chart, it is clear that we need a mix of campaign types. Otherwise, we will be successful in bringing customers to the site, but they won't convert. Or conversely, we will only be targeting those customers who have already been to our site, but will not bring in any new prospects.
- So, if we choose the 5 campaigns with the highest number of customers for each scenario (meaning those are the campaigns that reached the most people), we will know which ones CoolTShirts should re-invest in. See the highlighted rows below for my choices.

campaign	landing_page	shopping_cart	checkout	purchase	Total
cool-tshirts-search	171	133	7	2	313
getting-to-know-cool-tshirts	617	682	41	9	1349
interview-with-cool-tshirts-founder	625	515	31	7	1178
paid-search	-	-	179	52	231
retargeting-ad	-	-	445	113	558
retargeting-campaign	-	-	246	54	300
ten-crazy-cool-tshirts-facts	587	570	32	9	1198
weekly-newsletter	-	-	450	115	565

3.4 Optimizing the Campaign Budget

- In Conclusion, I chose the campaigns below to re-invest in, because they reached the most current and potential customers out of the eight total campaigns. By choosing these, we will be able to optimize the campaign budget and (hopefully) bring in more potential customers, as well as converting more existing customers.

campaign	landing_page	shopping_cart	checkout	purchase	Total
getting-to-know-cool-tshirts	617	682	41	9	1349
interview-with-cool-tshirts-founder	625	515	31	7	1178
retargeting-ad	-	-	445	113	558
ten-crazy-cool-tshirts-facts	587	570	32	9	1198
weekly-newsletter	-	-	450	115	565