



Attribution Capstone

Learn SQL from Scratch

Jeremiah Schupbach

18 July 2018

Project Table of Contents

1. Getting familiar with CoolTShirts
2. What is the user journey?
3. Optimizing the campaign budget

1. Getting familiar with Cool T-Shirts

1.1 How many campaigns and sources are there?

This was accomplished in two separate, simple queries (shown on right)

- Utm_campaign (Campaign_Total in below table) refers to a specific advertisement or email campaign
- Utm_source (Source_Total in below table) refers to which site sent the traffic, or where the campaign was placed

Campaign_Total	Source_Total
8	6

```
SELECT COUNT(DISTINCT utm_campaign) as  
'Campaign_Total'  
FROM page_visits;
```

```
SELECT COUNT(DISTINCT utm_source) as 'Source_Total'  
FROM page_visits;
```

1.2 How are the campaigns and sources related?

The query shown below was used to find distinct campaigns and their sources (results on the right). This matches the results from the previous slide as email and google each have two campaigns.

```
SELECT DISTINCT utm_campaign as 'Distinct_Campaign',  
               utm_source as 'Source'  
FROM page_visits;
```

Distinct_Campaign	Distinct_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

1.3 What pages are on the Cool T-Shirts site?

The query shown below was used to find all of the distinct pages on the Cool T-Shirts site, with the results shown on the right.

```
SELECT DISTINCT page_name as 'Distinct_Pages'  
FROM page_visits;
```

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

2. The User Journey

2.1 First Touches

First touches tell us how the user discovered the site by showing where they came from. The query shown on the right was used to find how many first touches each campaign is responsible for, with the results shown below.

Source	Campaign	Total
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 as 'Source',  
       ft_attr.utm_campaign as 'Campaign',  
       COUNT(*) as 'Total'  
FROM ft_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```


2.2 Last Touches

Last touches only consider the source on the most recent site visit for a user. The query shown on the right was used to find how many last touches each campaign is responsible for, with the results shown below.

Source	Campaign	Total
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

```
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 as 'Source',  
       lt_attr.utm_campaign as 'Campaign',  
       COUNT(*) as 'Total'  
FROM lt_attr  
GROUP BY 2  
ORDER BY 3 DESC;
```

2.3 User Purchases

While First and Last Touch are important to know, we also want to know how many users actually spent some money. The query used is in the grey box and will list how many users hit each page, with the results below it.

```
SELECT page_name as 'Site_Page',  
       count(distinct user_id) as 'Total_Users'  
FROM page_visits  
GROUP BY 1;
```

Site_Page	Total_Users
1 - landing_page	1979
2 - shopping_cart	1881
3 - checkout	1431
4 - purchase	361

2.4 Last Touch and Purchases

Knowing about the number of customers purchasing is great, but yields us little if we cannot link those purchases to specific campaigns. The query used (below right) will give us the results we seek (below left).

Source	Campaign	Total_P
email	weekly-newsletter	115
facebook	retargetting-ad	113
email	retargetting-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 as 'Source',  
       lt_attr.utm_campaign as 'Campaign',  
       COUNT(*) as 'Total_P'  
FROM lt_attr  
GROUP BY 2  
ORDER BY 3 DESC;
```

2.5 What is the typical user journey?

- Users typically enter the site via media source. This is based on only 8% of total first touches being for users googling specifically for t-shirts.
- 72% of users entering the site shopped, placed something in the cart, and made it to the checkout page.
- Of the users that made it to checkout, 25% made a purchase.

3. Optimizing the Budget

3.1 Campaign Re-Investment

- Re-invest in Medium, NYTimes, and BuzzFeed campaigns. With 92% of first touches driven by 3 media campaigns, the site must re-invest these top 3 first touch producers. Without this traffic, stable or increasing conversion rates won't matter.
- Re-invest in the newsletter email campaign and the facebook re-targeting. These Top 2 in the Last Touch + Purchase metric account for 63% of all purchases.
- The fact that re-targeting campaigns in general make up 78% of purchases seems to support the idea that keeping the first touch numbers as high as possible should be given more weight in this consideration. We cannot re-target people who do not show up in the first place.