



# Capstone Project

Learn SQL from Scratch

Kate Prater

8/11/2018



# Table of Contents

---



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

# **1. Get familiar with CoolTShirts**

# Get Familiar with CoolTShirts

*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`.*

- A campaign is the tool, ad, or promotion that influences someone to visit a page.
- A source is the channel through which the traffic was sent, e.g Google, email.
- By running queries to find distinct counts, we find that there are 8 different campaigns and 6 different sources.

```
SELECT Count(DISTINCT utm_campaign) AS 'Campaign Count'  
FROM page_visits;
```

```
SELECT Count(DISTINCT utm_source) AS 'Source Count'  
FROM page_visits;
```

Campaign Count
8

Source Count
6

# Get Familiar with CoolTShirts

- Each page visit is connected to a campaign and a source.
- In this dataset, some campaigns are utilizing the same source.
- It is possible that multiple sources could run the same campaign, but this is not represented in the current dataset.



```
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

# Get Familiar with CoolTShirts

*What pages are on their website?*

- By setting up the query to find distinct page names, we can see that CoolTShirts has a four-page path: landing page, shopping cart, checkout, and purchase.

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

```
SELECT DISTINCT page_name AS '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?*

- We must first find the minimum timestamp for the user to determine when they first visited a page.
- We can then attribute that time stamp to the campaign and source that prompted that first visit.

First Touch Source	First Touch 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 AS 'First Touch Source',  
       ft_attr.utm_campaign AS 'First Touch Campaign',  
       COUNT(*) AS '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?*

- Using a similar method as the first touch, we can attribute the last touch campaigns and sources by finding the maximum timestamps.

Last Touch Source	Last Touch 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 AS 'Last Touch Source',  
       lt_attr.utm_campaign AS 'Last Touch Campaign',  
       COUNT(*) AS 'Count'  
FROM lt_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

# What is the user journey?

## *How many visitors make a purchase?*

- We can find how many visitors made a purchase by completing a distinct count using a where clause or by grouping by page name.
- By using the 'Group By' method, we can see that about 18% of users who visited the landing page ended up making a purchase.

Count of Users Who Made a Purchase
361

Page Name	Count of Users
1 - landing_page	1979
2 - shopping_cart	1881
3 - checkout	1431
4 - purchase	361

```
SELECT COUNT(DISTINCT user_id) AS 'Count of Users Who Made a Purchase'
FROM page_visits
WHERE page_name = '4 - purchase';
```

--or

```
SELECT page_name AS 'Page Name', COUNT(DISTINCT user_id) AS 'Count of Users'
FROM page_visits
GROUP BY page_name;
```



# What is the user journey?

Last Touch Source	Last Touch Campaign	Count
email	weekly-newsletter	115
facebook	retargeting-ad	113
email	retargeting-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

```
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 'Last Touch Source',  
       lt_attr.utm_campaign AS 'Last Touch Campaign',  
       COUNT(*) AS 'Count'  
FROM lt_attr  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

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

- By using the maximum timestamp for the purchase page by user and linking the attributed last touch campaign, we are able to find a count of purchase page last touches by campaign. We can also connect the source to that last touch.

# What is the *typical* user journey?

## *What is the typical user journey?*

- The three campaigns that bring in the most users do not result directly to the most purchases (only 7%).
- In conjunction with those top “first touch campaigns”, the retargetting and weekly newsletter campaigns help bring users back and lead to the most purchases (78%).



Campaign	First Touch Count	Last Touch Count	Purchase Count
getting-to-know-cool-tshirts	612	232	9
weekly-newsletter	n/a	447	115
ten-crazy-cool-tshirts-facts	576	190	9
retargetting-campaign	n/a	245	54
retargetting-ad	n/a	443	113
interview-with-cool-tshirts-founder	622	184	7
paid-search	n/a	178	52
cool-tshirts-search	169	60	2

### **3. Optimize the campaign budget**

# Optimize the campaign budget

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

- Though they do not independently result in the most purchases, the three campaigns that bring in the most users are very important investments that ultimately contribute to purchases: getting-to-know-cool-tshirts, ten-crazy-cool-tshirts-facts, and interview-with-cool-tshirts-founder
- I would also suggest reinvesting in the two campaigns that result in the highest number of purchases: weekly-newsletter and retargeting-ad. Twenty-six percent of the last touches for both of these campaigns result in purchases.

