



# Capstone: Attribution

Learn SQL from Scratch

Brandon Steili

July 2018

# Table of Contents

1. Get familiar with CoolTShirts – Slides 3-6
2. What is the user journey? – Slides 7-12
3. Optimize the campaign budget – Slides 13-14

# **Getting Familiar with CoolTShirts**

# What's the difference between a campaign and a source?

A campaign is a group or series of advertisements that address a single idea or a single theme. For example if I wanted to sell widgets and I had a 4<sup>th</sup> of July themed widget, I would create an advertising campaign showing my stars and stripes themed widget. I would then target that campaign to run in the weeks prior to the 4<sup>th</sup> of July holiday to maximize potential sales of that specific widget.

A source in this context is the place a campaign was run on – like Facebook or Twitter which lead someone to take action on a particular campaign.

# 1.1 How Many Campaigns and Sources?

CoolTShirts uses **8 campaigns**:

```
SELECT count(DISTINCT utm_campaign) as campaign_count  
FROM page_visits;
```

campaign_count
8

CoolTShirts uses **6 sources**:

```
SELECT count(DISTINCT utm_source) as source_count  
FROM page_visits;
```

source_count
6

Each Campaign is limited to a single source, but they do run multiple campaigns on some sources:

```
SELECT utm_campaign, utm_source  
FROM page_visits  
GROUP BY utm_campaign;
```

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

# What Pages are on the CoolTshirts Website?

- There are 4 pages listed on the website each part of the purchase pipeline:

```
SELECT DISTINCT page_name as 'Page Name'  
FROM page_visits;
```

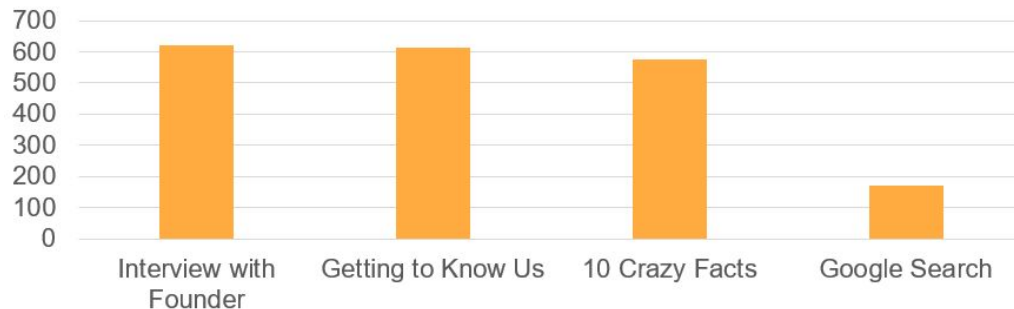
Query Results	
Page Name	
1 -	landing_page
2 -	shopping_cart
3 -	checkout
4 -	purchase

**What is the user journey?**

# How many first touches is each campaign responsible for?

The data here is surprisingly consistent across three of the campaigns, although the fourth (search) lags far behind the others and is not a significant source of first touches indicating that there may be some issues with brand awareness.

Query Results		
Source	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



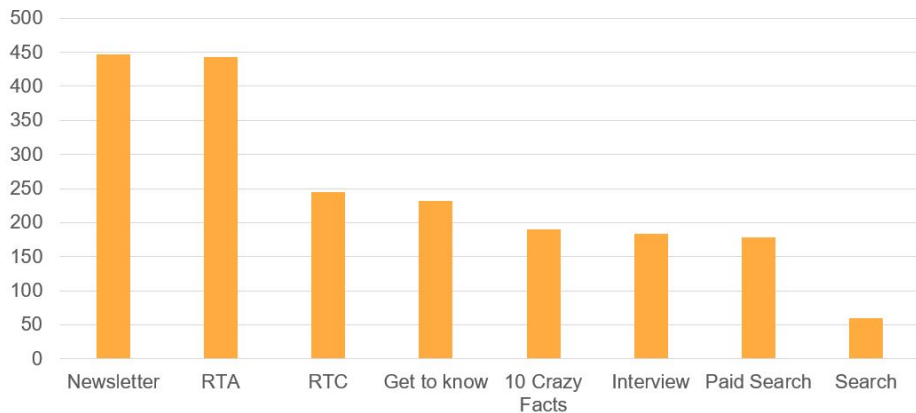
```
WITH first_touch AS (  
  SELECT user_id,  
         MIN(timestamp) as first_touch_at  
  FROM page_visits  
  GROUP BY user_id),  
first_touch_sandc AS (  
  SELECT ftouch.user_id,  
         ftouch.first_touch_at,  
         pvisit.utm_source,  
         pvisit.utm_campaign  
  FROM first_touch AS ftouch  
  JOIN page_visits AS pvisit  
  ON ftouch.user_id = pvisit.user_id  
  AND ftouch.first_touch_at = pvisit.timestamp  
)  
SELECT first_touch_sandc.utm_source as 'Source',  
       first_touch_sandc.utm_campaign as 'Campaign',  
       COUNT(*) as 'Count'  
FROM first_touch_sandc  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```



# How many last touches is each campaign responsible for?

The data here is much more inconsistent across the various methods. The weekly newsletter and Facebook retargeting are both well ahead of the other methods and individually almost double the returns from Google.

Query Results		
Source	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 last_touch AS (  
    SELECT user_id,  
           MAX(timestamp) as last_touch_at  
    FROM page_visits  
    GROUP BY user_id),  
last_touch_sandc AS (  
    SELECT ltouch.user_id,  
           ltouch.last_touch_at,  
           pvisit.utm_source,  
           pvisit.utm_campaign  
    FROM last_touch AS ltouch  
    JOIN page_visits AS pvisit  
    ON ltouch.user_id = pvisit.user_id  
    AND ltouch.last_touch_at = pvisit.timestamp  
)  
SELECT last_touch_sandc.utm_source as 'Source',  
       last_touch_sandc.utm_campaign as 'Campaign',  
       COUNT(*) as 'Count'  
FROM last_touch_sandc  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

# How many visitors make a purchase?

Out of 1979 distinct visitors CoolTshirts has converted 361 into sales or roughly 18%.

```
SELECT COUNT(DISTINCT user_id) as Purchasers  
FROM page_visits  
WHERE page_name is '4 - purchase';
```

Query Results	
	Purchasers
	361

Purchases vs. Visits

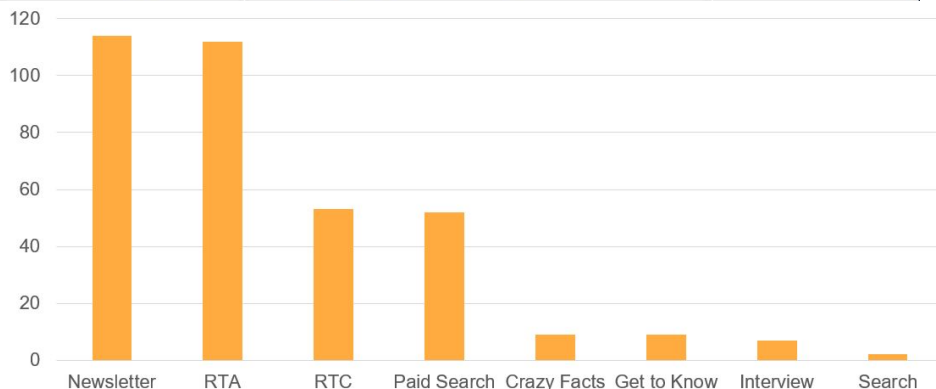


■ Visitors ■ Purchasers

# How many last touch purchases is each campaign responsible for?

Once again the data shows a couple of clear winning campaigns. Facebook and the weekly email campaign bring back more visitors and covert them to sales.

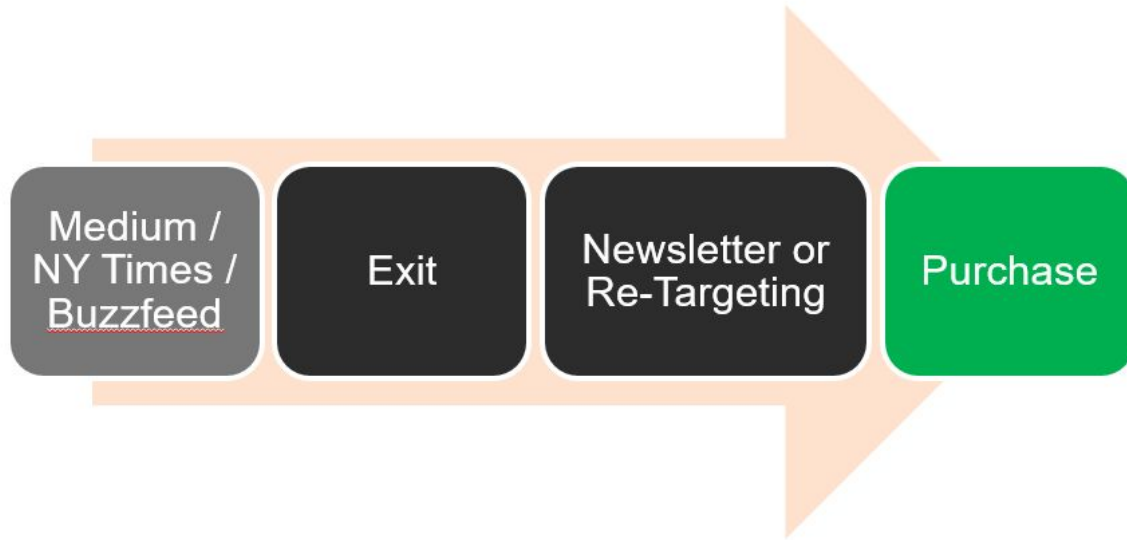
Query Results		
Source	Campaign	Count
email	weekly-newsletter	114
facebook	retargeting-ad	112
email	retargeting-campaign	53
google	paid-search	52
buzzfeed	ten-crazy-cool-tshirts-facts	9
nytimes	getting-to-know-cool-tshirts	9
medium	interview-with-cool-tshirts-founder	7



```
WITH last_touch AS (  
    SELECT user_id,  
           MAX(timestamp) as last_touch_at  
    FROM page_visits  
    GROUP BY user_id),  
last_touch_sandc AS (  
    SELECT ltouch.user_id,  
           ltouch.last_touch_at,  
           pvisit.utm_source,  
           pvisit.utm_campaign,  
           pvisit.page_name  
    FROM last_touch AS ltouch  
    JOIN page_visits AS pvisit  
    ON ltouch.user_id = pvisit.user_id  
    AND ltouch.last_touch_at = pvisit.timestamp  
)  
SELECT last_touch_sandc.utm_source as 'Source',  
       last_touch_sandc.utm_campaign as  
'Campaign',  
       COUNT(*) as 'Count'  
FROM last_touch_sandc  
WHERE last_touch_sandc.page_name is '4 - purchase'  
GROUP BY 1, 2  
ORDER BY 3 DESC;
```

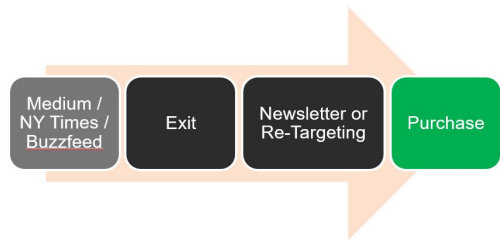
# What is the typical user journey?

Based upon the data the typical flow looks like this:



**Optimize the Budget**

# Optimize the Budget



- 1,810 first touch users reached CoolTShirts from three articles on three websites. That's 91.4% of our total visitors from three sources and three campaigns.
- CoolTShirts had 358 total purchases. 226 of those purchases or 63.1% were the results of two additional campaigns.

Based upon these numbers the recommendation would be to fund the following five campaigns due to their return on investment:

Interview with CoolTShirts founder – 31.4% of visitors

Getting to know CoolTShirts – 30.9% of visitors

Ten crazy CoolTShirts facts – 29.1% of visitors

Weekly Newsletter – 31.8% of purchases

Re-Targeting Ad – 31.2% of purchases