



Capstone: First and Last Touch Attribution

Learn SQL from Scratch

Anaanya Elizabeth Bijay

10/12/2018

Table of Contents

1. 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**.
- What pages are on their website?

2. What is the user journey?

- How many first touches is each campaign responsible for?
- How many last touches is each campaign responsible for?
- How many visitors make a purchase?
- How many last touches on the ***purchase page*** is each campaign responsible for?
- What is the typical user journey?

3. Optimize the campaign budget

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

1. Get familiar with CoolTShirts

1. 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**.

utm_source:

A source is the origin of your website's traffic, such as a search engine like Google.

A source can have multiple campaigns running.

Hence, utm_source is a custom variable that allows you to identify **where** the visits sent to a particular page came from.

utm_campaign:

A campaign is a set of activities used to promote a product.

A single campaign can be run across multiple sources.

Hence, utm_campaign is a custom variable that allows you to track **which** campaigns are sending visits to a particular page.

1. 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`.

```
1  /*1.1.1  number of distinct campaigns*/
2
3  SELECT COUNT(DISTINCT utm_campaign)
4  FROM page_visits;
5
6  /*1.1.2  number of distinct sources*/
7  SELECT COUNT(DISTINCT utm_source)
8  FROM page_visits;
9
10 /*1.1.3  campaigns and their related sources*/
11 SELECT DISTINCT utm_campaign, utm_source
12 FROM page_visits;
```

Query Results	
COUNT(DISTINCT utm_campaign)	
8	
COUNT(DISTINCT utm_source)	
6	
utm_campaign	utm_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

CoolTShirts runs 8 campaigns across 6 sources. It runs 2 campaigns on email and 2 search campaigns on Google.

None of the campaigns are repeated across the various sources.

1. Get familiar with CoolTShirts

What pages are on the CoolTShirts website? Find the distinct values of the **page_name** column.

```
1  /*1.2 Distinct pages on CoolTShirts website*/
2  SELECT DISTINCT page_name
3  FROM page_visits;
```

CoolTShirts has 4 pages on its website based on the page_visits table.

Interestingly, there does not seem to be any sub-categorisation of the landing page by product or category.

Hence, we can presume that the customer lands on a particular subset of the main landing page from the campaign link they were directed from.

Query Results
page_name
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

2. What is the user journey?

2. What is the user journey?

How many first touches is each campaign responsible for?

```
1  /* 2.1 count first touches by campaign*/
2  WITH first_touch AS (
3      SELECT user_id,
4             MIN(timestamp) as first_touch_at
5      FROM page_visits
6      GROUP BY user_id),
7  ft_attr AS (
8      SELECT ft.user_id,
9             ft.first_touch_at,
10            pv.utm_source,
11            pv.utm_campaign
12      FROM first_touch ft
13      JOIN page_visits pv
14          ON ft.user_id = pv.user_id
15          AND ft.first_touch_at = pv.timestamp
16  )
17  SELECT ft_attr.utm_source,
18         ft_attr.utm_campaign,
19         COUNT(*)
20  FROM ft_attr
21  GROUP BY 1, 2
22  ORDER BY 3 DESC;
```

Query Results		
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

While CoolTShirts is running 8 campaigns across 6 platforms, only 4 campaigns across 4 distinct platforms have resulted in first touches, with a total of 1979 first touches.

The “interview-with-cool-tshirts-founder” campaign on medium resulted in 622 (31.4%) first touches, followed by “getting-to-know-cool-tshirts” on nytimes (612 or 30.9%), “ten-crazy-cool-tshirts-facts” on buzzfeed (576 or 29.1%) and “cool-tshirts-search” on google with 169 first touches (8.54%).

2. What is the user journey?

How many last touches is each campaign responsible for?

```
1  /* 2.2 count last touches by campaign*/
2  WITH last_touch AS (
3      SELECT user_id,
4             MAX(timestamp) as last_touch_at
5      FROM page_visits
6      GROUP BY user_id),
7  lt_attr AS (
8      SELECT lt.user_id,
9             lt.last_touch_at,
10            pv.utm_source,
11            pv.utm_campaign,
12            pv.page_name
13      FROM last_touch lt
14      JOIN page_visits pv
15        ON lt.user_id = pv.user_id
16        AND lt.last_touch_at = pv.timestamp
17  )
18  SELECT lt_attr.utm_source,
19         lt_attr.utm_campaign,
20         COUNT(*)
21  FROM lt_attr
22  GROUP BY 1, 2
23  ORDER BY 3 DESC;
```

Query Results		
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

All 8 campaigns across 6 distinct platforms have resulted in last touches, with a total of 1979 last touches.

Conversely, the campaigns responsible for the most first touches where responsible for the least last touches. The top three campaigns responsible for last touches (“weekly-newsletter”, “retargetting-ad” and “retargetting-campaign”) were not responsible for a single first touch attribution.

2. What is the user journey?

How many visitors make a purchase?

```
1  /* 2.3 count last touches by campaign*/  
2  SELECT COUNT(DISTINCT user_id)  
3  FROM page_visits  
4  WHERE page_name = '4 - purchase';
```

Query Results
COUNT(DISTINCT user_id)
361

There are 361 users who made a purchase. Hence the campaigns have resulted in an overall conversion rate of 18.2%.

2. What is the user journey?

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

```
1  /* 2.4 count last touches on purchase page by
   campaign*/
2  WITH last_touch AS (
3      SELECT user_id,
4             MAX(timestamp) as last_touch_at
5      FROM page_visits
6      WHERE page_name = '4 - purchase'
7      GROUP BY user_id),
8  lt_attr AS (
9      SELECT lt.user_id,
10             lt.last_touch_at,
11             pv.utm_source,
12             pv.utm_campaign,
13             pv.page_name
14      FROM last_touch lt
15      JOIN page_visits pv
16          ON lt.user_id = pv.user_id
17          AND lt.last_touch_at = pv.timestamp
18  )
19  SELECT lt_attr.utm_source,
20         lt_attr.utm_campaign,
21         COUNT(*)
22  FROM lt_attr
23  GROUP BY 1, 2
24  ORDER BY 3 DESC;
```

Query Results		
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

All 8 campaigns across the 6 distinct sources can be attributed to the last touches on the purchase page.

The campaign with the most last touch attributions overall and on the purchase page is the “weekly-newsletter” email campaign and the campaign with the least last touch attributions overall and on the purchase page is the “cool-tshirts-search” google campaign.

The purchase page results closely resemble the overall last touch attribution results.

2. What is the user journey?

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

```
1  /* 2.4 count last touches on purchase page by
   campaign*/
2  WITH last_touch AS (
3      SELECT user_id,
4              MAX(timestamp) as last_touch_at
5      FROM page_visits
6      WHERE page_name = '4 - purchase'
7      GROUP BY user_id),
8  lt_attr AS (
9      SELECT lt.user_id,
10             lt.last_touch_at,
11             pv.utm_source,
12             pv.utm_campaign,
13             pv.page_name
14     FROM last_touch lt
15     JOIN page_visits pv
16         ON lt.user_id = pv.user_id
17         AND lt.last_touch_at = pv.timestamp
18 )
19 SELECT lt_attr.utm_source,
20         lt_attr.utm_campaign,
21         COUNT(*)
22 FROM lt_attr
23 GROUP BY 1, 2
24 ORDER BY 3 DESC;
```

Query Results		
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

All 8 campaigns across the 6 distinct sources can be attributed to the last touches on the purchase page.

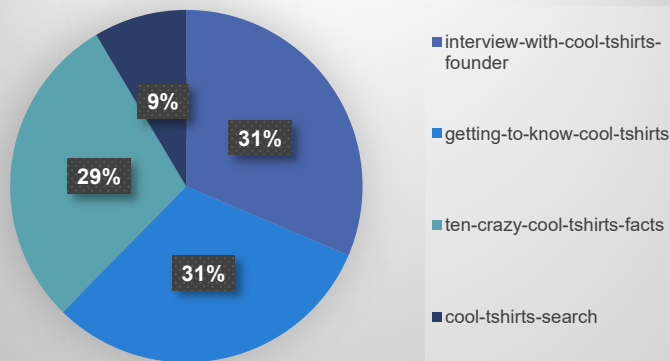
The campaign with the most last touch attributions overall and on the purchase page is the “weekly-newsletter” email campaign and the campaign with the least last touch attributions overall and on the purchase page is the “cool-tshirts-search” google campaign.

The purchase page results closely resemble the overall last touch attribution results.

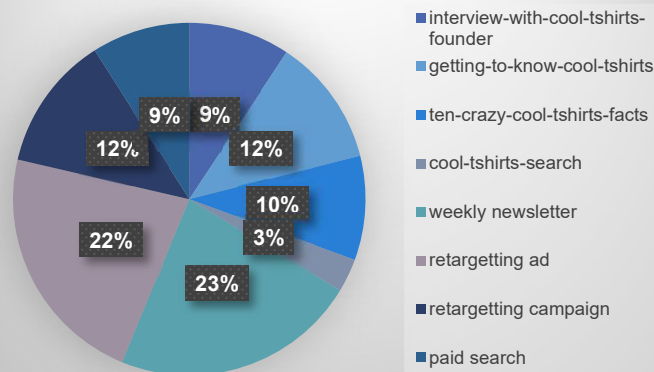
2. What is the user journey?

What is the typical user journey?

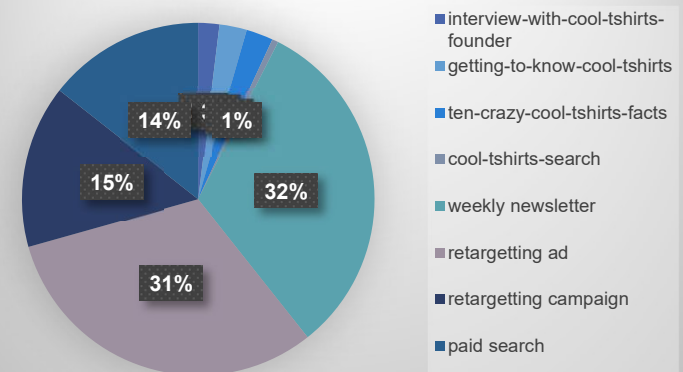
First Touch by Campaign



Last Touch by Campaign



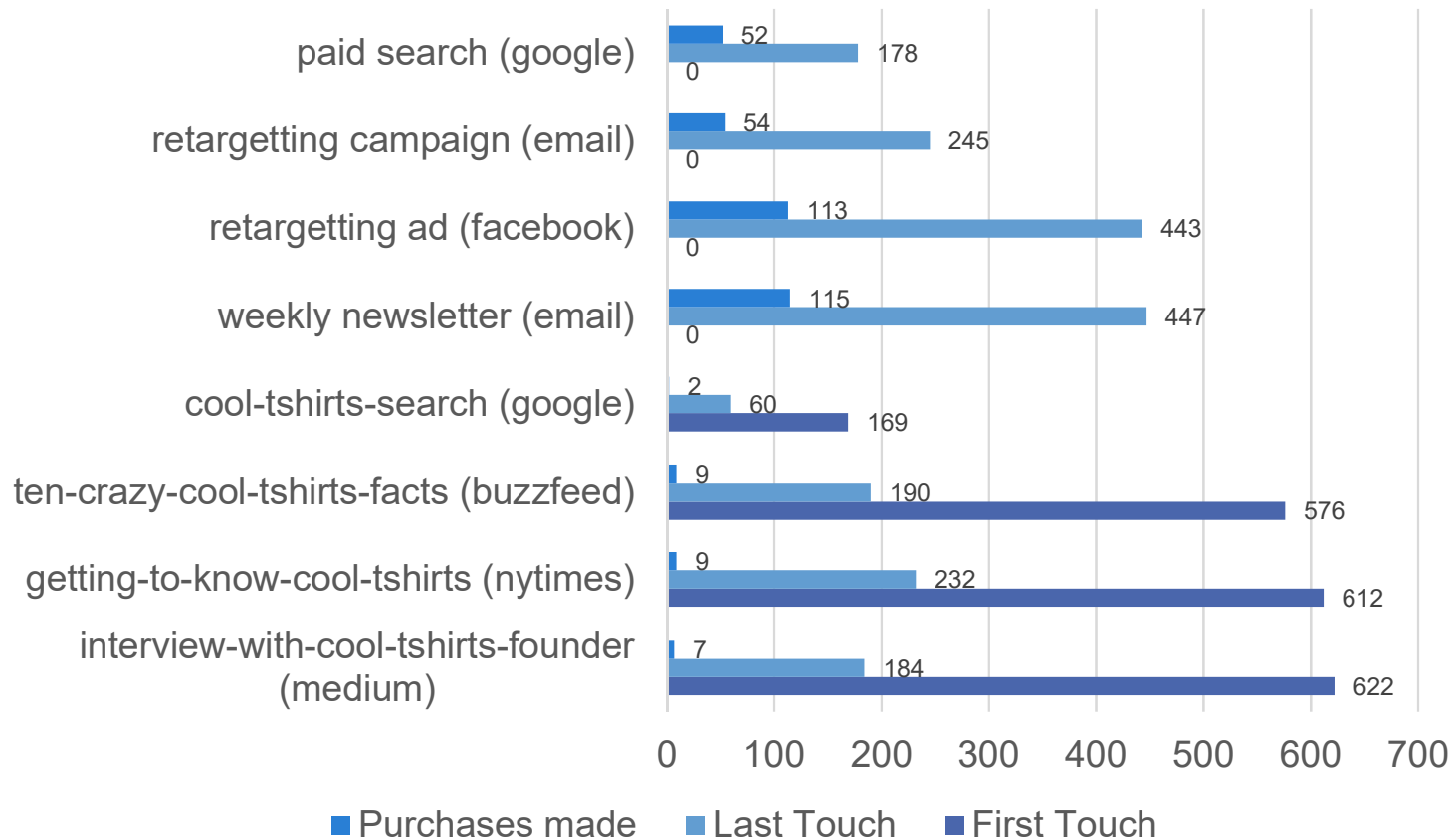
Purchases by Campaign



2. What is the user journey?

What is the typical user journey?

Typical user journey



2. What is the user journey?

What is the typical user journey?

As seen in the charts above, 91% of first touches originated from “content-based” marketing campaigns (articles & stories) with near equal first touches originating from the medium and nytimes.

However, most customers needed a reminder about the products as only 6% of purchases could be attributed to the three “content-based” marketing campaigns, with 78% of the purchases being attributed to “reminder-based” campaigns (newsletters & retargeting ads).

Hence, the typical user journey can thus be summed up as firstly, an interaction with a “content-based” marketing campaign, secondly, an interaction with a “reminder-based” campaign followed by a purchase.

3. Optimize the campaign budget

3. Optimize the campaign budget

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

Campaign	First Touch	First Touch (%)	Last Touch	Last Touch (%)	Purchases	Purchases (%)
interview-with-cool-tshirts-founder (medium)	622	31%	184	9%	7	2%
getting-to-know-cool-tshirts (nytimes)	612	31%	232	12%	9	2%
ten-crazy-cool-tshirts-facts (buzzfeed)	576	29%	190	10%	9	2%
cool-tshirts-search (google)	169	9%	60	3%	2	1%
weekly newsletter (email)	0	0%	447	23%	115	32%
retargetting ad (facebook)	0	0%	443	22%	113	31%
retargetting campaign (email)	0	0%	245	12%	54	15%
paid search (google)	0	0%	178	9%	52	14%

3. Optimize the campaign budget

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

As seen in the chart above, all the campaigns have had purchases attributed to them.

However, as the campaigns that generated the most first touches are not the same as the campaigns that generated the most sales, a balanced approach must be taken when prioritising the campaigns to reinvest.

The three “content-based” marketing campaigns can be attributed with 91% of first-touches, indicating that this is an effective strategy to generate customer interest.

Conversely, the three “reminder-based” campaigns cannot be attributed to a single first touch but make up 78% of purchases, indicating that while they may not be the best campaigns to generate customer buzz, most customers respond better to a reminder-based strategy.

The remaining “search-based” campaign type should be discontinued, as only 9% of first touches can be attributed to such campaigns even though they account for 14% of purchases. This indicates that the medium is largely organically driven and hence, may not need as much investment to as brand awareness is needed to drive up “search-based” purchases.

3. Optimize the campaign budget

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

Hence, CoolTShirts should proceed with a blend of the 2 strategies and reinvest in a mix of both “content-based” and “reminder-based” campaigns as it would generate the most interest and sales.

Of the 3 “content-based” and 3 “reminder-based” campaigns, CoolTShirts should continue to invest in all 3 “content-based” campaigns as it is an effective way to reach out to new audiences.

As for the reminder-based campaigns, CoolTShirts should continue with the 2 retargeting campaigns as they can be spread across a wide range of sources including on the same page as the “content-based” campaign articles as it would help the company to maximise exposure and sales via the same channel.

The company could choose not to reinvest in the weekly newsletter articles as it reaches out to an existing set of customers across email, which while is a medium that generates significant sales, does not generate any new interest.

3. Optimize the campaign budget

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

Hence, CoolTShirts should re-invest in “interview-with-cool-tshirts-founder”, “getting-to-know-cool-tshirts”, “ten-crazy-cool-tshirts-facts”, “retargeting ad” and “retargeting campaign”.