



First- and Last-Touch Attribution with CoolTShirts.com

Learn SQL from Scratch—Capstone Project

Eric Timmons

July 20, 2018

Table of Contents

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

1. Getting Familiar with CoolTShirts

1.1 Campaign Count

Campaign Count = 8

- Unique campaigns and number of website page visits for each advertising campaign are shown below in the query results
- Users are identified by which advertising campaign they clicked to reach one of the web site pages.
- Three campaigns stand out and are highlighted below.

utm_campaign	Campaign Cnt
getting-to-know-cool-tshirts	1349
ten-crazy-cool-tshirts-facts	1198
interview-with-cool-tshirts-founder	1178
weekly-newsletter	565
retargeting-ad	558
cool-tshirts-search	313
retargeting-campaign	300
paid-search	231

```
/* Query to get a complete campaign count for all campaigns */  
SELECT COUNT(DISTINCT utm_campaign) AS 'Campaign Count'  
FROM page_visits;
```

```
/* Query to get an idea of campaigns relative to page visits */  
SELECT utm_campaign, COUNT (utm_source) AS 'Campaign Cnt'  
FROM page_visits pv  
GROUP BY utm_campaign  
ORDER BY COUNT(utm_source) DESC;
```

1.2 Source Count

Source Count = 6

- Unique Sources and number of page visits for each Source are shown below in the query
- Sources are the location where the advertisements were run

utm_source	Source Cnt
buzzfeed	1198
email	865
facebook	558
google	544
medium	1178
nytimes	1349

```
/* Query to get a distinct source count */  
SELECT COUNT(DISTINCT utm_source) AS 'Source Count'  
FROM page_visits;
```

```
/* Query to evaluate distribution of page visits by source */  
SELECT DISTINCT utm_source, COUNT(utm_campaign) AS 'Source  
Cnt'  
FROM page_visits  
GROUP BY utm_source  
;
```

1.2 Source and Campaign Relationship

- Campaigns appear to be run relative to source.
- The SQL Query attached illustrates for each campaign, which source it was targeted to.
- Each campaign run has one source, however multiple campaigns could be targeted to a single source. (e.g. Google and email)
- *It would be instructive to see how campaigns applied to multiple sources would perform.*

utm_campaign	SUM (buzz)	SUM(email)	SUM(fb)	SUM(google)	SUM(med)	SUM(nyt)
cool-tshirts-search	0	0	0	313	0	0
getting-to-know-cool-tshirts	0	0	0	0	0	1349
interview-with-cool-tshirts-founder	0	0	0	0	1178	0
paid-search	0	0	0	231	0	0
retargeting-ad	0	0	558	0	0	0
retargeting-campaign	0	300	0	0	0	0
ten-crazy-cool-tshirts-facts	1198	0	0	0	0	0
weekly-newsletter	0	565	0	0	0	0

```
WITH sources AS(
SELECT utm_source, utm_campaign,
CASE
  WHEN utm_source LIKE 'buzzfeed'
  THEN 1 ELSE 0 END as 'buzz',
CASE
  WHEN utm_source LIKE 'email'
  THEN 1 ELSE 0 END as 'email',
CASE
  WHEN utm_source LIKE 'facebook'
  THEN 1 ELSE 0 END as 'fb',
CASE
  WHEN utm_source LIKE 'google'
  THEN 1 ELSE 0 END as 'google',
CASE
  WHEN utm_source LIKE 'medium'
  THEN 1 ELSE 0 END as 'med',
CASE
  WHEN utm_source LIKE 'nytimes'
  THEN 1 ELSE 0 END as 'nyt'
FROM page_visits
)
SELECT utm_campaign, SUM (buzz), SUM(email), SUM(fb),
SUM(google),
SUM(med), SUM(nyt)
FROM sources
GROUP BY utm_campaign;
```

1.3 Pages on the Website

Page names visited on the website are listed below and are a progression from Landing Page to Purchase.

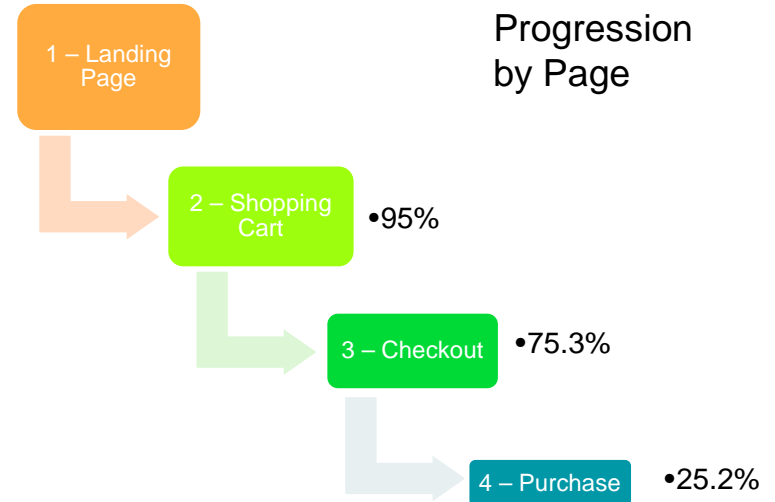
As with a classic funnel structure, the volume decreases with each progressive step.

- 95% go from Landing Page to Shopping Cart
- 75.3% go from Shopping Cart to Checkout
- 25.2% go from Checkout to Purchase

With the large number of users abandoning at checkout, it would be instructive to review the page to determine if an issue and to ensure they are targeted in the retargeting campaigns or create a special one.

page_name	COUNT (page_name)
1 - landing_page	2000
2 - shopping_cart	1900
3 - checkout	1431
4 - purchase	361

```
/* Distinct Page names and their frequency of occurrence */  
SELECT page_name, COUNT (page_name)  
FROM page_visits  
GROUP BY page_name  
ORDER BY page_name;
```



1.4 Unique Users

Unique users in the data set count 1,979.

With 361 purchasers, it represents an 18.2% close rate overall.

```
SELECT COUNT(DISTINCT user_id)
FROM page_visits;
```


2. User Journey

2.1 First Touch Count by campaign

Campaigns with First Touch attribution

Of the 8 campaigns, only 4 have first touch.

- The other 4 are paid search, retargeting and weekly newsletter, campaigns likely designed to increase purchase rates after initial prospect acquisition.

For the campaigns, the interview has a slight lead for first touch.

utm_campaign	First Touch Count
interview-with-cool-tshirts-founder	622
getting-to-know-cool-tshirts	612
ten-crazy-cool-tshirts-facts	576
cool-tshirts-search	169

```
/* First Touch by campaign */
WITH first_touch AS (
  SELECT user_id,
         MIN(timestamp) as first_touch_at
  FROM page_visits
  GROUP BY user_id)
SELECT pv.utm_campaign,
       COUNT(pv.utm_source) AS 'First Touch Count'
FROM first_touch ft
JOIN page_visits pv
  ON ft.user_id = pv.user_id
  AND ft.first_touch_at = pv.timestamp
GROUP BY pv.utm_campaign
;
```

2.2 Last Touch Count by campaign

Campaigns with Last Touch attribution

For the campaigns, the weeklynewsletter has a slight lead for last touch followed closely by retargeting-ad.

Weekly-newsletter, retargeting-ad and retargeting-campaign are designed to reinforce the advertising message and keep in touch with potential clients, which appears to be working as they are the top 3 Last touch campaigns.

utm_campaign	Last Touch 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

```
/* Last Touch by campaign */
WITH last_touch AS (
  SELECT user_id,
         MAX(timestamp) as last_touch_at
  FROM page_visits
  GROUP BY user_id)
SELECT pv.utm_campaign,
       COUNT(pv.utm_source) AS 'Last Touch Count'
FROM last_touch lt
JOIN page_visits pv
  ON lt.user_id = pv.user_id
 AND lt.last_touch_at = pv.timestamp
GROUP BY pv.utm_campaign
ORDER BY COUNT(pv.utm_source) DESC
;
```

2.3 Purchases by Campaign (Last Touch Attribution)

Purchasers, those visiting the purchase page were 361 (Query attached)

Purchases based on last attribution are shown below (Totals 358)

NOTE: See next page for discussion on the difference between last attribution and purchase page

Important to note that purchases are mainly driven by retargeting not the initial campaign.

utm_campaign	Purchase Count
weekly-newsletter	114
retargeting-ad	112
retargeting-campaign	53
paid-search	52
getting-to-know-cool-tshirts	9
ten-crazy-cool-tshirts-facts	9
interview-with-cool-tshirts-founder	7
cool-tshirts-search	2

```
SELECT COUNT(DISTINCT user_id) AS 'Unique Purchasers'
FROM page_visits
WHERE page_name LIKE '4 - purchase';
```

```
/* Last Touch with Purchase by campaign */
WITH last_touch AS (
  SELECT user_id,
         MAX(timestamp) as last_touch_at
  FROM page_visits
  GROUP BY user_id)
SELECT pv.utm_campaign,
       COUNT(pv.utm_source) AS 'Purchase Count'
FROM last_touch lt
JOIN page_visits pv
  ON lt.user_id = pv.user_id
  AND lt.last_touch_at = pv.timestamp
WHERE pv.page_name LIKE '4 - purchase'
GROUP BY pv.utm_campaign
ORDER BY COUNT(pv.utm_source) DESC
;
```

2.4 Progression and Last Touch Attribution Note

In looking at the last touch, it was noticed that 358 had a last touch for page 4-Purchase, however in looking at the number of users that had visited the 4Purchase page, the total was 361.

In researching, the three users that represent the difference had visited another page after purchase, typically arriving from a different source.

This is a small amount for this data set, but should be watched as customers that may browse for other items might be missed. A more complex query would be needed to used to identify purchases with subsequent page visits since LT alone would not.

```
SELECT * FROM page_visits
WHERE user_id = 94852
ORDER BY timestamp DESC;
```

```
SELECT * FROM page_visits
WHERE user_id = 29180;
```

```
SELECT * FROM page_visits
WHERE user_id = 76466;
```

page_name	timestamp	user_id	utm_campaign	utm_source
2 - shopping_cart	2018-01-27 07:19:09	94852	getting-to-know-cool-tshirts	nytimes
1 - landing_page	2018-01-27 02:35:09	94852	getting-to-know-cool-tshirts	nytimes
4 - purchase	2018-01-13 09:07:45	94852	retargetting-ad	facebook
3 - checkout	2018-01-13 08:57:45	94852	retargetting-ad	facebook
2 - shopping_cart	2018-01-09 03:10:45	94852	ten-crazy-cool-tshirts-facts	buzzfeed
1 - landing_page	2018-01-09 02:41:45	94852	ten-crazy-cool-tshirts-facts	buzzfeed

page_name	timestamp	user_id	utm_campaign	utm_source
2 - shopping_cart	2018-01-28 13:45:24	29180	getting-to-know-cool-tshirts	nytimes
1 - landing_page	2018-01-28 13:20:24	29180	getting-to-know-cool-tshirts	nytimes
4 - purchase	2018-01-12 16:36:41	29180	retargetting-campaign	email
3 - checkout	2018-01-12 16:22:41	29180	retargetting-campaign	email
2 - shopping_cart	2018-01-10 02:43:41	29180	getting-to-know-cool-tshirts	nytimes
1 - landing_page	2018-01-10 00:29:41	29180	getting-to-know-cool-tshirts	nytimes

page_name	timestamp	user_id	utm_campaign	utm_source
2 - shopping_cart	2018-01-12 22:08:05	76466	interview-with-cool-tshirts-founder	medium
1 - landing_page	2018-01-12 22:02:05	76466	interview-with-cool-tshirts-founder	medium
4 - purchase	2018-01-10 15:39:09	76466	weekly-newsletter	email
3 - checkout	2018-01-10 15:35:09	76466	weekly-newsletter	email
2 - shopping_cart	2018-01-09 06:28:09	76466	ten-crazy-cool-tshirts-facts	buzzfeed
1 - landing_page	2018-01-09 01:33:09	76466	ten-crazy-cool-tshirts-facts	buzzfeed

2.5 Purchasers Only by First Touch by Campaign

The attached query illustrates the First Touch attribution for only those that purchased (361). The idea being to determine if those that purchase have a different distribution than the overall population visting the web site.

The order does not appear changed from the first touch distribution for all users.

utm_campaign	First Touch	Percent
interview-with-cool-tshirts-founder	118	32.69
ten-crazy-cool-tshirts-facts	109	30.19
getting-to-know-cool-tshirts	103	28.53
cool-tshirts-search	31	8.59

```
/* Purchasers First Touch Attribution */
WITH purch_cam AS (
WITH purchasers AS (
  SELECT user_id
  FROM page_visits
  WHERE page_name LIKE '4 - purchase'),
  total_purchases AS (
    SELECT COUNT(DISTINCT user_id) AS 'Unique Purchasers'
    FROM purchasers
  )
SELECT p.user_id,
                                pv.utm_campaign, pv.utm_source,
min(pv.timestamp), pv.page_name,
  total_purchases.'Unique Purchasers' AS Total
FROM purchasers p
JOIN page_visits pv
  ON p.user_id = pv.user_id
CROSS JOIN total_purchases
GROUP BY p.user_id
ORDER BY p.user_id ASC, pv.timestamp DESC
)
SELECT pc.utm_campaign, COUNT(pc.user_id) as "First Touch",
ROUND(100.0*COUNT(pc.user_id)/pc.Total,2) AS 'Percent'
FROM purch_cam pc
GROUP BY pc.utm_campaign
ORDER BY COUNT(pc.user_id) DESC
;
```

2.6 Purchasers Only by Last Touch by Campaign

The attached query illustrates the Last Touch attribution for only those that purchased (361)

The order does not appear changed from the Last Touch distribution in general.

utm_campaign	Last Touch	Percent
weekly-newsletter	114	31.58
retargeting-ad	112	31.02
retargeting-campaign	53	14.68
paid-search	52	14.4
getting-to-know-cool-tshirts	11	3.05
ten-crazy-cool-tshirts-facts	9	2.49
interview-with-cool-tshirts-founder	8	2.22
cool-tshirts-search	2	0.55

```
WITH purch_campaign AS (  
  WITH purchasers AS (  
    SELECT user_id  
    FROM page_visits  
    WHERE page_name LIKE '4 - purchase'  
  ),  
  total_purchases AS (  
    SELECT COUNT(DISTINCT user_id) AS 'Unique Purchasers'  
    FROM purchasers  
  )  
  SELECT p.user_id, pv.utm_campaign,  
         pv.utm_source, max(pv.timestamp), pv.page_name,  
         total_purchases.'Unique Purchasers' AS Total  
  FROM purchasers p  
  JOIN page_visits pv  
    ON p.user_id = pv.user_id  
  CROSS JOIN total_purchases  
  GROUP BY p.user_id  
  ORDER BY p.user_id ASC, pv.timestamp DESC  
)  
SELECT pc.utm_campaign, COUNT(pc.user_id) as "Last Touch",  
       ROUND(100.0*COUNT(pc.user_id)/pc.Total,2) AS 'Percent'  
FROM purch_campaign pc  
GROUP BY pc.utm_campaign  
ORDER BY COUNT(pc.user_id) DESC;
```

2.7 Purchasers from Retargetting Campaigns Origins

The attached query and results were run to see if the retargeting campaigns (weekly-newsletter, retargetting-ad, retargetting-campaign) have a different distribution based on source, ie was one source significantly better than another.

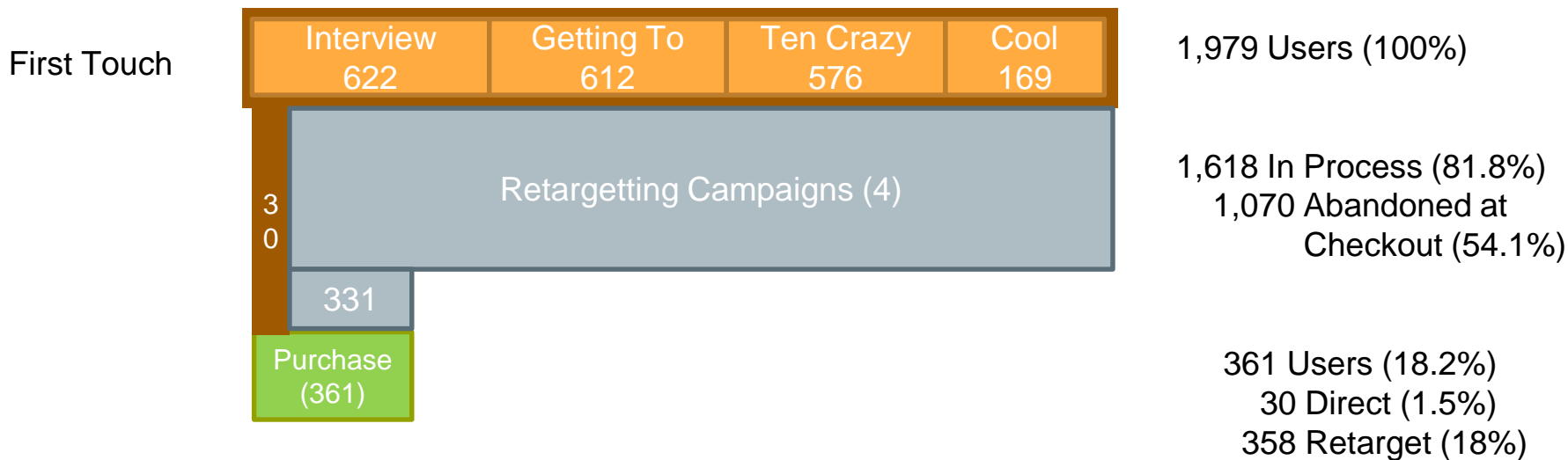
The results indicate that purchasers purchasing as a result of a retargetting campaign were evenly distributed across the primary three First Touch campaigns. No significant weighting different from FT data. Interview still the strongest.

FT Campaign	Retargetted Purchasers
cool-tshirts-search	29
getting-to-know-cool-tshirts	94
interview-with-cool-tshirts-founder	111
ten-crazy-cool-tshirts-facts	100

```
/* Purchasers last touch = retargetting source campaign (i.e. First Touch) */
WITH purchasers AS (
  SELECT user_id
  FROM page_visits
  WHERE page_name LIKE '4 - purchase'
),
last_touch AS (
  SELECT p.user_id, pv.utm_campaign, pv.utm_source,
    max(pv.timestamp)
  FROM purchasers p
  JOIN page_visits pv
    ON (p.user_id = pv.user_id
      AND (pv.utm_campaign LIKE 'weekly%' OR pv.utm_campaign
        LIKE 'retargetting%'
          OR pv.utm_campaign LIKE 'paid%'))
  GROUP BY p.user_id
),
first_touch AS (
  SELECT p.user_id, pv.utm_campaign, pv.utm_source,
    min(pv.timestamp)
  FROM purchasers p
  JOIN page_visits pv
    ON p.user_id = pv.user_id
  JOIN last_touch lt
    ON p.user_id = lt.user_id
  GROUP BY p.user_id
)
SELECT ft.utm_campaign AS 'FT Campaign', COUNT(ft.user_id) AS 'Retargetted
Purchasers'
FROM first_touch ft
GROUP BY ft.utm_campaign;
```


3. Optimizing the Campaign Budget

3.1 Pipeline Analysis



The retargeting campaigns are essential to securing purchases on the website. Of the 361 purchasers on the site, 331 (92.5%) came from the retargeting campaigns (4) demonstrating the need to keep retargeting efforts strong.

3.2 Campaigns to Reinvest

- Initial Customer Acquisition Campaigns for reinvestment
 1. getting-to-know-cool-tshirts
 2. ten-crazy-cool-tshirts-facts
 3. interview-with-cool-tshirts-founder

These customer acquisition campaigns had the highest First Touch attribution, ranging from 576 to 622. The search campaign came in far underneath.

- Retargeting Ads
 1. Weekly Newsletter
 2. Retargeting-Ad

Retargeting Ads are essential to gaining closure on the sale. The two campaigns listed had the highest attribution for purchasers 63% of purchasers (226/361) were from these two campaigns.

3.3 Other Items to Consider

- Campaigns are directly related to sources. It would be instructive to run campaigns across multiple sources and review correlation between campaigns and sources to see the impact of sources.
- Retargeting campaigns are essential. A number of users (1070 or 54%) exited at Checkout without finalizing the purchase.
 - Retargeting campaigns to remind users they had items in their cart would be useful. Since these users are easily identified in a SQL Query, they could be targeted with a unique retargeting campaign.