codecademy

SQL Capstone: Attribution

Learn SQL from Scratch
Kevin Neff
August 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 – Campaigns and Sources

How many campaigns and sources does CoolTShirts use? What is the difference between a campaign and source?

- A **source** is a type of communication, typically email, or website on which a campaign is exposed to prospective customers.
- A **campaign** is a specific message that is intended to elicit a response from a specific group of potential customers.
- It is important to note that a source can run multiple campaigns and vice versa. For example one source could run a newsletter campaign and a sale campaign.
- To get the number of distinct campaigns, run a SELECT COUNT(DISTINCT) query. Also, we can label the column names by including as. These queries are shown on the next slide.

Distinct Campaigns	Distinct Sources
8	6

Campaigns	Sources
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 – Pages

What pages are on the CoolTShirts Website?
- The best way to find the page names is to run a SELECT DISTINCT query, as this shows us the *unique* page names.
- Each of the names refers to a different part of the CoolTShirts website.

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

```
1   /*
2   These queries show us the number of distinct campaigns, as
    well as the sources per campaign.
3   */
4   SELECT COUNT(DISTINCT utm_campaign) AS 'Campaign Count'
5   FROM page_visits;
6
7   SELECT COUNT (DISTINCT utm_source) AS 'Source Count'
8   FROM page_visits;
9
10  SELECT COUNT(DISTINCT utm_campaign) AS 'Distinct Campaigns',
    COUNT(DISTINCT utm_source) AS 'Distinct Source'
11  FROM page_visits;
12
13  SELECT DISTINCT utm_campaign AS 'Campaign'
14  , utm_source AS 'Source'
15  FROM page_visits;
16
17  SELECT DISTINCT page_name AS 'Page Name'
18  FROM page_visits;
```

2. What is the User Journey?

What is the User Journey? First Touch

How many *first* touches is each campaign responsible for?
- MIN(timestamp) returns the *first* touch information for each user.
- There are nearly 2,000 unique visitors, but only four campaigns were responsible for the first touch.

Below: Beginning of first touch query, showing use of MIN(timestamp).

```
WITH first_touch AS (
    SELECT user_id,
        MIN(timestamp) as first_touch_at
    FROM page_visits
    GROUP BY user_id),
```

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

What is the User Journey? Last Touch

How many *last* touches is each campaign responsible for?

- MAX(timestamp) returns the *last* touch information for each user.
- Last touch is important because it shows the last source and campaign that drove the clients last visit to the site.

Below: Beginning of the last touch query, which utilizes MAX(timestamp)

```
42  WITH last_touch AS (
43      SELECT user_id,
44          MAX(timestamp) as last_touch_at
45      FROM page_visits
46      GROUP BY user_id),
```

Source	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

Queries: First Touch vs. Last Touch

```sql
20    /*First Touch Attribution Query*/
21
22    WITH first_touch AS (
23        SELECT user_id,
24            MIN(timestamp) as first_touch_at
25        FROM page_visits
26        GROUP BY user_id),
27    ft_attr AS(
28      SELECT ft.user_id,
29            ft.first_touch_at,
30            pv.utm_source,
31            pv.utm_campaign
32      FROM first_touch ft
33      JOIN page_visits pv
34        ON ft.user_id = pv.user_id
35        AND ft.first_touch_at = pv.timestamp
36    )
37    SELECT ft_attr.utm_source AS Source,
38          ft_attr.utm_campaign AS Campaign,
39          COUNT (*) AS COUNT
40    FROM ft_attr
41    GROUP BY 1, 2
42    ORDER BY 3 DESC;
```

Left: First Touch Attribution Query

Right: Last Touch Attribution Query

```sql
44    /*Last Touch Attribution Query*/
45
46    WITH last_touch AS (
47        SELECT user_id,
48            MAX(timestamp) as last_touch_at
49        FROM page_visits
50        GROUP BY user_id),
51    lt_attr AS (
52    SELECT lt.user_id,
53          lt.last_touch_at,
54          pv.utm_source,
55          pv.utm_campaign,
56          pv.page_name
57    FROM last_touch lt
58    JOIN page_visits pv
59      ON lt.user_id = pv.user_id
60      AND lt.last_touch_at = pv.timestamp
61      )
62      SELECT lt_attr.utm_source AS Source,
63            lt_attr.utm_campaign AS Campaign,
64            COUNT(*)
65      FROM lt_attr
66      GROUP by 1, 2
67      ORDER BY 3 DESC;
```

What is the User Journey? Conversions and Purchases

How many visitors make a purchase?
- A simple query with WHERE (*below*) shows us that 361 users made it to the purchase page.
- Conversion Rate is 18.2%

How many last touches *on the purchase page* is each campaign responsible for?
- The table below shows how many last touches each campaign is responsible for.
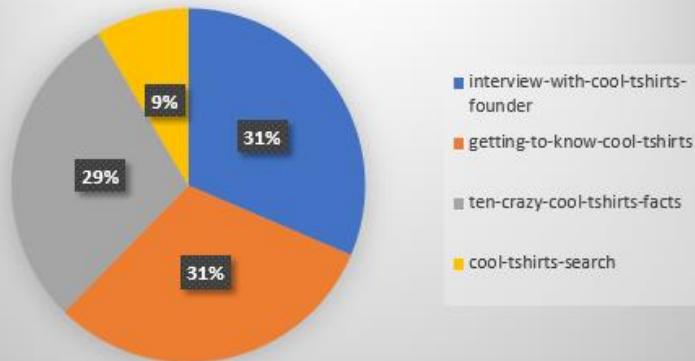- MAX(timestamp) and a WHERE clause help us determine which campaigns were most effective.

```sql
SELECT COUNT(DISTINCT user_id) AS 'Customers'
FROM page_visits
WHERE page_name = '4 - purchase';
```

Source	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

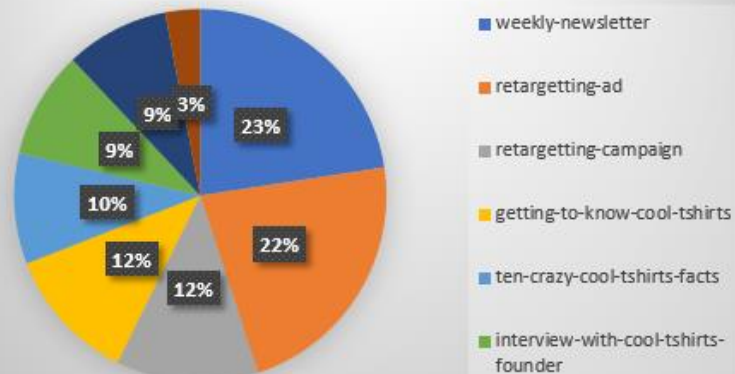# What is the TYPICAL User Journey? - Touches by the Numbers

- Current event campaigns, such as "Interview with Cool T Shirt Founder" grab peoples' attention. Approximately 91% of first touches are from these types of campaigns

- Customers ultimately end up clicking through to the website via other campaigns, such as the retargeting campaigns.
- 57% of last touches can be attributed to retargeting.

First Touches By Campaign

- 31% — interview-with-cool-tshirts-founder
- 31% — getting-to-know-cool-tshirts
- 29% — ten-crazy-cool-tshirts-facts
- 9% — cool-tshirts-search

Last Touch by Campaign

- 23% — weekly-newsletter
- 22% — retargetting-ad
- 12% — retargetting-campaign
- 12% — getting-to-know-cool-tshirts
- 10% — ten-crazy-cool-tshirts-facts
- 9% — interview-with-cool-tshirts-founder
- 9%
- 3%

What is the TYPICAL User Journey? – Purchases by the Numbers

- Retargeting campaigns might not have initially gotten the customer's attention, but they did close the sale. 78.1% of purchases can be attributed to the retargeting campaigns and the weekly newsletter.
- The campagins with strong first touch numbers have incredibly weak numbers here. Together they account for about 8% of purchases, with the remaining 14% being attributed to Paid Search.
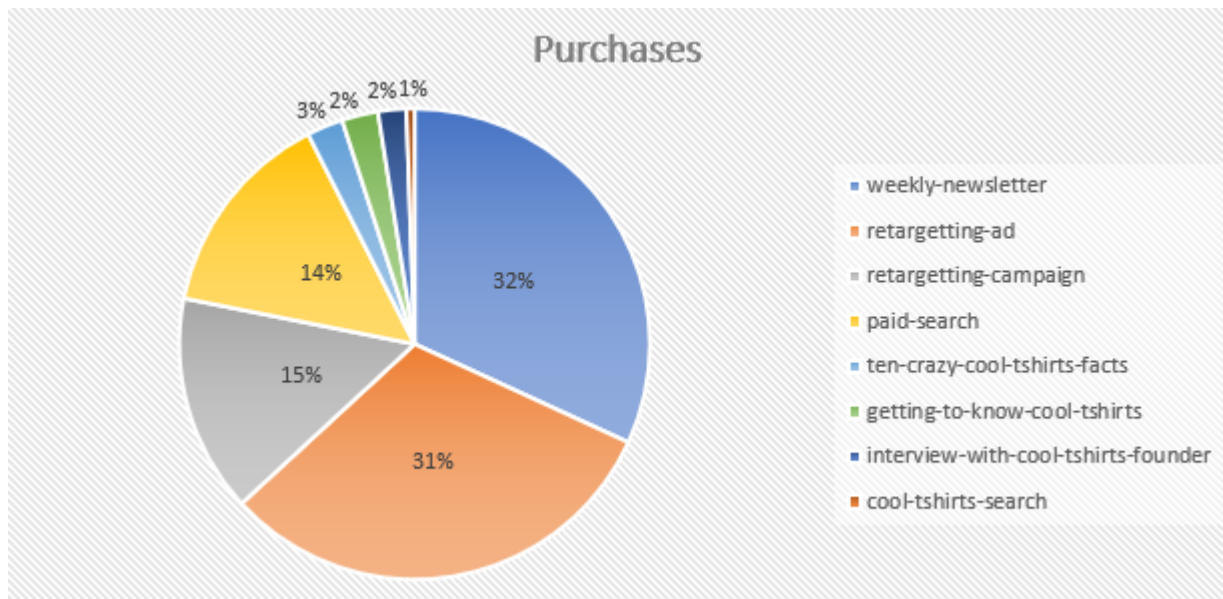
3. Optimizations

Optimizing Budget – Top 5 Campaigns to Leverage

- The "Current" campaigns are effective at getting potential customers on the site, while the retargeting campaigns are all deliver most of the purchases.
  - Likewise, it makes sense to invest ad spend in a blend of these campaigns.

- As "getting people in door" is important, it makes sense to use all three campaigns of the Current campaigns. Based off of first touch rates, these three current campaigns are best campaigns of their type to use:
  - Getting to know CoolTShirts
  - Interview with CoolTShirts Founder
  - Ten Crazy CoolTShirts facts

- Retargeting is highly effective for closing the sale. Based off of purchase page visits, these Retargeting/ Newsletter campaigns are the best of their type to optimize towards:
  - Retargeting Ad
  - Weekly Newsletter

- A table showing the data for each campaign is on the next slide. The highlighted campaigns are the campaigns that are most effective.

Optimizing Budget – Top 5 Campaigns to Leverage

Campaign	# of First Touches	% of First Touches	# of Last Touches	% of Last Touches	# of Purchases	% of Purchases
cool-tshirts-search	169	8.5%	60	3.0%	2	0.6%
getting-to-know-cool-tshirts	612	**30.9%**	232	11.7%	9	2.5%
interview-with-cool-tshirts-founder	622	**31.4%**	184	9.3%	7	1.9%
paid-search	0	0.0%	178	9.0%	52	14.4%
retargetting-ad	0	0.0%	443	22.4%	113	**31.3%**
retargetting-campaign	0	0.0%	245	12.4%	54	15.0%
ten-crazy-cool-tshirts-facts	576	**29.1%**	190	9.6%	9	2.5%
weekly-newsletter	0	0.0%	447	22.6%	115	**31.9%**
ALL CAMPAIGNS	*1979*	*100.0%*	*1979*	*100.0%*	*361*	*100.0%*