Note!

This Powerpoint goes into fairly granular detail about the SQL code (for eg. why certain commands are used, and when). I did this deliberately to force myself to write out my thought process in a step-by-step manner and is a way for me to learn the ins-and-outs of SQL at a much deeper level.

codecademy

CoolTShirts: First- and Last-Touch Attribution Analysis

Richmond Wong

5 June 2018 Cohort

Table of Contents

Section 1
Getting Familiar with the Company

Getting Familiar With The Company

How many campaigns and sources does CoolTShirts use and how are they related?

There are a total of 8 distinct campaigns (a mix of email-based and website-based campaigns). The websites that CoolTShirts is running campaigns from is Buzzfeed, Facebook, Google, Medium and The New York Times.

There are a total of 6 distinct campaign sources.

What is the difference between utm_campaign and utm_source?

utm_campaigns are the marketing campaigns that are run from CoolTShirts' email list and website-based campaigns (some are obviously paid such as those on Google and NYTimes while others may be free to run – such as the interview with the company founder on Medium (where anyone can post content for free)).

utm_source are the websites that the *utm_campaigns* are run from, and also include two email campaigns (*retargetting-campaign and weekly-newsletter*).

SQL CODE:

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

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

What Pages Are On The Site?

What are the names of the pages on the site?

There are a total of 4 distinct pages: *1- landing_page, 2 - shopping_cart, 3 - checkout* and *4 - purchase*.

One interesting note is that the email campaigns link to the purchase page only and not to any of the other pages.

SQL CODE:

```
SELECT distinct page_name, utm_campaign,
utm_source
FROM page_visits
GROUP BY page_name, utm_campaign
ORDER BY page_name DESC;
```

page_name	utm_campaign	utm_source
1 - landing_page	cool-tshirts-search	google
1 - landing_page	getting-to-know-cool-tshirts	nytimes
1 - landing_page	interview-with-cool-tshirts-founder	medium
1 - landing_page	ten-crazy-cool-tshirts-facts	buzzfeed
2 - shopping_cart	cool-tshirts-search	google
2 - shopping_cart	getting-to-know-cool-tshirts	nytimes
2 - shopping_cart	interview-with-cool-tshirts-founder	medium
2 - shopping_cart	ten-crazy-cool-tshirts-facts	buzzfeed
3 - checkout	cool-tshirts-search	google
3 - checkout	getting-to-know-cool-tshirts	nytimes
3 - checkout	interview-with-cool-tshirts-founder	medium
3 - checkout	paid-search	google
3 - checkout	retargetting-ad	facebook
3 - checkout	retargetting-campaign	email
3 - checkout	ten-crazy-cool-tshirts-facts	buzzfeed
3 - checkout	weekly-newsletter	email
4 - purchase	cool-tshirts-search	google
4 - purchase	getting-to-know-cool-tshirts	nytimes
4 - purchase	interview-with-cool-tshirts-founder	medium
4 - purchase	paid-search	google
4 - purchase	retargetting-ad	facebook
4 - purchase	retargetting-campaign	email
4 - purchase	ten-crazy-cool-tshirts-facts	buzzfeed
4 - purchase	weekly-newsletter	email

Section 2
The User Journey

User Journey – Step 1 of How Many First Touches Is Each Campaign Responsible For?

- In addition to the method taught in the First Touches lesson, we can also think about the question in the following way and still arrive at the same result:

  "For each distinct user, when was their First Touch and what is the campaign and utm_source associated with that First Touch?"

- In other words, we are first isolating for specific (ie. *distinct*) *user_ids* and their *min(timestamp)* only. Then we add two additional columns by SELECTing the *utm_campaign* and *utm_source* rows associated with these specific *user_id*s and *min(timestamp)* entries.

- This will give us a single row table. We need to GROUP BY *user_id* to break down this single row table into a multi-row table showing all *user_id*s and their *min(timestamp)*, *utm_campaign* and *utm_source*.

- With the chart that is compiled (shown right) we can move onto Step 2 (tallying the number of First Touches in chart form)

SQL CODE:

```
SELECT distinct user_id, min(timestamp), utm_campaign,
utm_source
FROM page_visits
GROUP BY user_id;
```

user_id	min(timestamp)	utm_campaign	utm_source
10006	2018-01-24 03:12:16	getting-to-know-cool-tshirts	nytimes
10030	2018-01-25 20:32:02	ten-crazy-cool-tshirts-facts	buzzfeed
10045	2018-01-05 18:31:17	getting-to-know-cool-tshirts	nytimes
10048	2018-01-16 04:17:46	interview-with-cool-tshirts-founder	medium
10069	2018-01-02 23:14:01	ten-crazy-cool-tshirts-facts	buzzfeed
10162	2018-01-29 21:37:10	getting-to-know-cool-tshirts	nytimes
10177	2018-01-24 07:10:33	getting-to-know-cool-tshirts	nytimes
10254	2018-01-23 22:27:18	interview-with-cool-tshirts-founder	medium

User Journey – Step 2 of How many First Touches is each campaign responsible for?

- In this second step of calculating First Touches, we first alias the previous table from the previous slide as *first_touch*.

- Then, we SELECT the *utm_campaign and utm_source* columns from the *first_touch* table. We also use a count(*) to tally up the number of rows associated with each *utm_campaign*. Basically, GROUPing BY *utm_campaign* allows us to break the campaign data from one row into its multiple constituent rows (the resulting table would only be one row if we did not use GROUP BY).

SQL CODE:

```
WITH first_touch AS
(SELECT distinct user_id, min(timestamp),
utm_campaign, utm_source
FROM page_visits
GROUP BY user_id)
SELECT first_touch.utm_campaign,
first_touch.utm_source, count(*)
FROM first_touch
GROUP BY 1, 2
ORDER BY 3 DESC;
```

utm_campaign	utm_source	count
interview-with-cool-tshirts-founder	medium	622
getting-to-know-cool-tshirts	nytimes	612
ten-crazy-cool-tshirts-facts	buzzfeed	576
cool-tshirts-search	google	169

User Journey - How many Last Touches is each campaign responsible for?

- We can calculate the number of Last Touches that each campaign is responsible for using SQL code similar to how we calculated First Touches in the previous two slides.

- The only difference in the SQL code is that we now SELECT for *max(timestamp)* rather than *min(timestamp)* to reflect the fact that we are interested in Last Touches.

SQL CODE:

```
WITH last_touch AS
(SELECT distinct user_id, max(timestamp),
utm_campaign, utm_source
FROM page_visits
GROUP BY user_id)
SELECT last_touch.utm_campaign,
last_touch.utm_source, count(*)
FROM last_touch
GROUP BY 1, 2
ORDER BY 3 DESC;
```

utm_campaign	utm_source	count(*)
weekly-newsletter	email	447
retargetting-ad	facebook	443
retargetting-campaign	email	245
getting-to-know-cool-tshirts	nytimes	232
ten-crazy-cool-tshirts-facts	buzzfeed	190
interview-with-cool-tshirts-founder	medium	184
paid-search	google	178
cool-tshirts-search	google	60

How many visitors make a purchase?

- This particular question makes an important assumption: that a visit to the purchase page itself constitutes a completed purchase. Also, the wording of the question appears to be asking us to find the number of users who have made at least one purchase (rather than asking us how many purchases in total were made).

- Accepting these assumptions, we can go onto the calculate the number of visitors who made a purchase by SELECTing and COUNTing the number of DISTINCT user_ids WHERE the *page_name* was equal to *4 – purchase*. This by itself will return a one row table showing the total number of unique visitors who made a purchase.

- We find that there are 361 unique users who made at least one purchase.

- We can then optionally SELECT *page_name* and *utm_campaign* to add columns to show the campaign name and page name (to make sure we're filtering for just *4 – purchase*). However, this by itself won't tell us anything about the number of unique buyers and the campaign that brought them to *4 – purchase* because the compiled table will still have one row only.

- To break down the total number of users who made at least one purchase by campaign, we need to include a GROUP BY *utm_campaign* in the SQL query. We will then get the table on the bottom breaking the 361 transactions into their *utm_campaigns*.

SQL CODE:

```
SELECT COUNT(DISTINCT user_id)
FROM page_visits
WHERE page_name = "4 - purchase";
```

count(distinct user_id)
361

SQL CODE:

```
SELECT COUNT(DISTINCT user_id) AS
"counter", page_name, utm_campaign
FROM page_visits
WHERE page_name = "4 - purchase"
GROUP BY utm_campaign
ORDER BY counter DESC;
```

counter	page_name	utm_campaign
115	4 - purchase	weekly-newsletter
113	4 - purchase	retargetting-ad
54	4 - purchase	retargetting-campaign
52	4 - purchase	paid-search
9	4 - purchase	getting-to-know-cool-tshirts
9	4 - purchase	ten-crazy-cool-tshirts-facts
7	4 - purchase	interview-with-cool-tshirts-founder
2	4 - purchase	cool-tshirts-search

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

- For this particular question, we start with all the *max(timestamp)* entries and the DISTINCT *user_ids* associated with them (each user will only have one Last Touch – ie. their most recent visit)

- Our next step is to filter only for the entries that meet the above requirements with respect to *max(timestamp)* and *user_ids* and also WHERE the page visited is *page_name = "4 – purchase"*.

- Having done this, we still only have a single row table. We need to GROUP BY *utm_campaign* so that we break down *utm_campaign* column into its constituent campaigns. This will then also show the COUNT for each of the individual campaigns (as shown on the chart on the right).

- We see that the table on this slide is exactly the same as the table on the last slide. We see that all purchases were from Last Touches. This makes sense as a First Touch would automatically be a Last Touch if a user bought via the first page they landed on.

SQL CODE:

```
SELECT COUNT(DISTINCT user_id) as count,
max(timestamp), page_name, utm_campaign, utm_source
FROM page_visits
WHERE page_name = "4 - purchase"
GROUP BY utm_campaign
ORDER BY count DESC;
```

counter	max(timestamp)	page_name	utm_campaign	utm_source
115	2018-02-04 11:09:47	4 - purchase	weekly-newsletter	email
113	2018-02-03 05:29:01	4 - purchase	retargetting-ad	facebook
54	2018-02-03 19:11:12	4 - purchase	retargetting-campaign	email
52	2018-02-04 00:17:45	4 - purchase	paid-search	google
9	2018-01-28 15:30:07	4 - purchase	getting-to-know-cool-tshirts	nytimes
9	2018-01-31 00:23:24	4 - purchase	ten-crazy-cool-tshirts-facts	buzzfeed
7	2018-01-20 17:18:05	4 - purchase	interview-with-cool-tshirts-founder	medium
2	2018-01-18 00:25:00	4 - purchase	cool-tshirts-search	google

Section 3
Additional Analysis & Optimizing the Campaign Budget – Five Campaigns to Reinvest In

Disproportionate Drop Off at Purchase Page – Examining the Funnel

Percentage of users who answer the next question

- We can see that there is an overly disproportionate drop off from checkout to purchase. In the third column of the table on this slide, only 25% of the users who made it to the checkout page then went on to complete their purchase.
- This doesn't make that much sense because 75% of users who put something in their shopping cart then decided to move to the checkout phase.
- We don't expect the percentage of users who move from each step of the funnel to be proportionate – but what we're experiencing here is an overly disproportionate decline given the percentages in the previous steps)

Percentage of users who complete the next step (using the original 2000 users as a base)

- It might be easier to appreciate the changes as we move through the 4 pages on the website if we use the original 2000 unique as a base. In the fourth column of the table we see that only 18% of the original 2000 users make a purchase, compared to 72% of the original 2000 users who make it all the way to the checkout.
- The sharp, overly disproportionate drop off from checkout to purchase should be looked at by the team as it indicates that something is turning users off from completing their purchase. This could be an overly complex purchase mechanism, or a buggy purchase interface / back-end for example. Either way, CoolTShirts.com needs to give this a good hard look.

SQL CODE:

```
SELECT page_name, COUNT(user_id)
FROM page_visits
GROUP BY page_name;
```

page_name	count(user_id)	% of users who answer the next question	% of users who complete each step (using original 2000 as base)
1 - landing_page	2000		
2 - shopping_cart	1900	95%	95%
3 - checkout	1431	75%	72%
4 - purchase	361	25%	18%

No Repeat Sales

- The SQL code on the top COUNTs the number of *user_id*s (including duplicates) that completed purchases. The code below it COUNTs the number of unique *user_id*s that completed purchases.

- Both produced exactly the same table, which indicates that there were no repeat purchases by any customer in the period we have data for.

- CoolTShirts.com will likely want to find out why customers are not coming back again for repeat business.

- Is it because the period we are measuring (1 Jan 2018 to 4 Feb 2018) too short? Or are there other reasons that none of the 361 customers returned to buy anything else?

SQL CODE:

```
SELECT COUNT(user_id) AS ”counter”,
utm_source
FROM page_visits
WHERE page_name = “4 – purchase”
GROUP BY utm_source
ORDER BY counter DESC;
```

SQL CODE:

```
SELECT COUNT(DISTINCT user_id) AS
“counter”, page_name, utm_campaign
FROM page_visits
WHERE page_name = “4 – purchase”
GROUP BY utm_campaign
ORDER BY counter DESC;
```

counter	page_name	utm_campaign
115	4 - purchase	weekly-newsletter
113	4 - purchase	retargetting-ad
54	4 - purchase	retargetting-campaign
52	4 - purchase	paid-search
9	4 - purchase	getting-to-know-cool-tshirts
9	4 - purchase	ten-crazy-cool-tshirts-facts
7	4 - purchase	interview-with-cool-tshirts-founder
2	4 - purchase	cool-tshirts-search

Only 6% of Visits Convert to a Sale

- The code at the top compiles the overall visits to the CoolTShirts.com website, including First Touches, Last Touches along with anything in between. There are a total of 5692 visits to the CoolTShirts.com website.
- The code at the bottom shows that there were 361 overall visits to the purchase page (including repeat visits by the same *user_id* – though in the previous slide we saw that there we no repeat visits to the purchase page)
- This means that only 6% of the total overall visits (First Touch, Last Touch and everything in between) converted to an actual sale.
- This seems like a fairly low number (for eg. Warby Parker in the Usage Funnels capstone is able to convert 25% of their users from initial contact using a survey to completing a purchase)

SQL CODE:

```
SELECT COUNT(user_id) AS ”counter”,
utm_source
FROM page_visits
GROUP BY utm_source
ORDER BY counter DESC;
```

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

SQL CODE:

```
SELECT COUNT(user_id) AS "counter",
utm_source
FROM page_visits
WHERE page_name = "4 - purchase"
GROUP BY utm_source
ORDER BY counter DESC;
```

counter	utm_source
169	email
113	facebook
54	google
9	buzzfeed
9	nytimes
7	medium

Optimize Campaign Budget – First Touches

- We see that the highlighted campaigns bring in the lion's share of First Touches (totaling 91% of all first visits!), driving awareness of the CoolTShirts website and brand. They were also responsible for a third (31%) of all Last Touches.

- However, these same campaigns are abysmal at converting these visits to actual sales (totaling just 6% of the 361 sales made in this period)

Campaign	Source	First Touch (FT) Count	FT %	Last Touch (LT) Count	LT %	Purchase Count	Purchase %
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%
Total		**1979**	**100%**	**1979**	**100%**	**361**	**100%**

Optimize Campaign Budget – Last Touches

- Three campaigns were responsible for 67% of Last Touches and together accounted for 78% of all sales!

- However, none of these campaigns were responsible for any First Touches, which might indicate that these were all prior visitors (ie. Those who already knew about CoolTShirts before the 1 Jan 2018 to 4 Feb 2018 period that this dataset encompasses).

Campaign	Source	First Touch (FT) Count	FT %	Last Touch (LT) Count	LT %	Purchase Count	Purchase %
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%
Total		**1979**	**100%**	**1979**	**100%**	**361**	**100%**

Optimize Campaign Budget – Reinvestment in Five Campaigns

If CoolTShirts was limited to just 5 campaigns to re-invest in, my suggestion are those highlighted in green below.

- Although *interview-with-cool-tshirts-founder*, *getting-to-know-cool-tshirts*, *ten-crazy-cool-tshirts-facts* barely drove any Last Touches and Purchases, they are critical to getting a steady stream of would-be customers to learn about the TCoolShirts site and brand. As mentioned before, they drive 91% of all First Touches!

- A significant chunk of purchases (63%) are derived from *weekly-newsletter*, *retargetting-ad* so CoolTShirts should also definitely reinvest in these campaigns.

Why not reinvest in targetting-campaign instead of ten-crazy-cool-tshirts-facts?

- Although *retargetting-campaign* (in red) contributes 15% to sales transactions, I would advise CoolTShirts to prioritize *ten-crazy-cool-tshirts-facts* over it because it's important to keep a steady flow of new customers (ie. "fresh blood) coming in.

- *crazy-cool-tshirts-facts* also contributes close to one-third (29%) of all First Touches so raising the number of would-be customers through greater investment in this campaign may subsequently result in a higher volume of sales transactions. Therefore, even if *weekly-newsletter* and *retargetting-ad* hold steady at 32% and 31% vis-a-vis total sales transactions, total revenue would have increased by investing in *ten-crazy-cool-tshirts* instead.

Campaign	Source	First Touch (FT) Count	FT %	Last Touch (LT) Count	LT %	Purchase Count	Purchase %
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%
Total		**1979**	**100%**	**1979**	**100%**	**361**	**100%**