



Capstone Project

Attribution Queries – Learning SQL from Scratch

Mafalda Patrício

12/11/18

Table of Contents

1. [Get familiar with CoolTShirts](#)
2. [What is the user journey?](#)
3. [Optimize the campaign budget](#)

1. Get Familiar with CoolTshirts

1.1 QUESTION

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.

- They have 8 campaigns and 6 sources. They are related in the way we can see through the figure b).
- An utm_campaign is the utm parameter (used by site owners to capture when and how a user finds a certain site) that identifies the specific ad or email blast used (i.e., ten-crazy-cool-tshirts-facts or weekly-newsletter), whereas the utm_source is an utm parameter that tells which site/touchpoint sent the traffic to the website (i.e., nytimes, email, buzzfeed or facebook).

a) N° of campaigns and sources:

COUNT(DISTINCT utm_campaign)
8
COUNT(DISTINCT utm_source)
6

b) How are campaigns and sources related?

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

1.2 QUESTION

What pages are on their website?

- They have 4 pages on their website: 1 – landing_page; 2 – shopping_cart; 3 – checkout; 4 – purchase. As can be seen through figure c).

QUERY USED:

```
SELECT DISTINCT page_name  
FROM page_visits;
```

c) N° of pages on the website

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

2. What is the User Journey?

2.1 QUESTIONS

How many first touches is each campaign responsible for? How many last touches is each campaign responsible for? How many visitors make a purchase?

- Through figures d) and e) we can see how many first and last touches each campaign is responsible for.
- Through figure f) we can see that 361 visitors made a purchase.

d) N° first touches each campaign is responsible for

ft_attr.utm_campaign	COUNT(*)
interview-with-cool-tshirts-founder	622
getting-to-know-cool-tshirts	612
ten-crazy-cool-tshirts-facts	576
cool-tshirts-search	169

e) N° last touches each campaign is responsible for

lt_attr.utm_campaign	COUNT(*)
weekly-newsletter	447
retargetting-ad	443
retargetting-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

f) N° visitors making a purchase

COUNT(DISTINCT user_id)
361

2.2 QUESTION

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

- Through figure g) we can see how many last touches on the purchase_page each campaign is responsible for.

g) N° last touches each campaign is responsible for at the purchase_page

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

2.3 QUESTION

What is the typical user journey?

- As we can see through the past analyzed figures and figures h) and i), out of the 1979 visitors in total of CoolTshirts website, the 3 campaigns that attract the most visitors are respectively: interview-with-cool-tshirts-founder, getting-to-know-cool-tshirts and ten-crazy-cool-tshirts-facts. These 3 campaigns out of the 4 in total used to attract initial visits, account for approx. 91% of all the visits the website has.

QUERY USED (for figure h):

```
SELECT COUNT(DISTINCT user_id),  
page_name  
FROM page_visits  
WHERE page_name = '1 -  
landing_page';
```

h) N° Visitors in total

COUNT(DISTINCT user_id)	page_name
1979	1 - landing_page

i) N° first touches each campaign is responsible

ft_attr.utm_source	ft_attr.utm_campaign	ft_attr.page_name	COUNT(*)
medium	interview-with-cool-tshirts-founder	1 - landing_page	622
nytimes	getting-to-know-cool-tshirts	1 - landing_page	612
buzzfeed	ten-crazy-cool-tshirts-facts	1 - landing_page	576
google	cool-tshirts-search	1 - landing_page	169

QUERY USED (for figure i):

```
WITH first_touch AS (  
  SELECT user_id,  
  MIN(timestamp) as first_touch_at  
  FROM page_visits  
  GROUP BY user_id),  
ft_attr AS (  
  SELECT ft.first_touch_at,  
  pv.utm_source,  
  pv.utm_campaign,  
  pv.timestamp,  
  pv.page_name  
  FROM first_touch AS 'ft'  
  JOIN page_visits AS 'pv'  
  ON ft.user_id = pv.user_id  
  AND ft.first_touch_at =  
  pv.timestamp)  
SELECT ft_attr.utm_source,  
ft_attr.utm_campaign,  
ft_attr.page_name, COUNT(*)  
FROM ft_attr  
GROUP BY 1, 2  
ORDER BY 4 DESC;
```

2.3 QUESTION

What is the typical user journey?

QUERY USED (for figure j):

```
WITH last_touch AS (  
  SELECT user_id,      MAX(timestamp)  
  as last_touch_at    FROM page_visits  
  GROUP BY user_id), lt_attr AS (SELECT  
  DISTINCT lt.user_id,  
  lt.last_touch_at, pv.utm_source,  
  pv.utm_campaign,  pv.timestamp,  
  pv.page_name  
  FROM last_touch AS 'lt'  
  JOIN page_visits AS 'pv'  
  ON lt.user_id = pv.user_id  
  AND lt.last_touch_at = pv.timestamp)  
SELECT lt_attr.utm_source,  
  lt_attr.utm_campaign,  
  lt_attr.page_name, COUNT(*)  
FROM lt_attr  
GROUP BY 1, 2, 3  
ORDER BY 2, 4 DESC;
```

j) N° last touches each campaign is responsible

utm_source	utm_campaign	page_name	COUNT(*)
google	cool-tshirts-search	2 - shopping_cart	46
google	cool-tshirts-search	1 - landing_page	7
google	cool-tshirts-search	3 - checkout	5
google	cool-tshirts-search	4 - purchase	2
nytimes	getting-to-know-cool-tshirts	2 - shopping_cart	153
nytimes	getting-to-know-cool-tshirts	1 - landing_page	38
nytimes	getting-to-know-cool-tshirts	3 - checkout	32
nytimes	getting-to-know-cool-tshirts	4 - purchase	9
medium	interview-with-cool-tshirts-founder	2 - shopping_cart	132
medium	interview-with-cool-tshirts-founder	3 - checkout	24
medium	interview-with-cool-tshirts-founder	1 - landing_page	21
medium	interview-with-cool-tshirts-founder	4 - purchase	7
google	paid-search	3 - checkout	126
google	paid-search	4 - purchase	52
facebook	retargetting-ad	3 - checkout	331
facebook	retargetting-ad	4 - purchase	112
email	retargetting-campaign	3 - checkout	192
email	retargetting-campaign	4 - purchase	53
buzzfeed	ten-crazy-cool-tshirts-facts	2 - shopping_cart	126
buzzfeed	ten-crazy-cool-tshirts-facts	1 - landing_page	32
buzzfeed	ten-crazy-cool-tshirts-facts	3 - checkout	23
buzzfeed	ten-crazy-cool-tshirts-facts	4 - purchase	9
email	weekly-newsletter	3 - checkout	333
email	weekly-newsletter	4 - purchase	114

2.3 QUESTION

What is the typical user journey?

- Through the previous analyzed figures and figure j) we can see that the campaigns that attract the most visitors initially to the webpage are not the same ones that make them engage the most in the act of the purchase.
- The campaigns that secure the most users effectively buying at CoolTshirts website are respectively: weekly-newsletter via e-mail (responsible for 114 of last touches), followed by retargeting-ad via facebook (responsible for 112 of last touches) and by paid-search via google (responsible for 52 of last touches).
- By conclusion, all this data shows us that the CoolTshirts typical user journey starts by initially attracting visitors mainly via news websites and later securing them as customers via targeted and specific adds through social media, e-mail or google.

3. Optimize the Campaign Budget

3. QUESTION

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

- Based on all the data and figures previously studied, CoolTshirts should definitely re-invest in the following 5 campaigns: interview-cool-tshirts-founder, getting-to-know-cool-tshirts, ten-crazy-cool-tshirts-facts, weekly-newsletter and retargeting-ad.
- The company should re-invest in the 3 first suggested campaigns as they are the ones to attract the most websites visit as a first-touch-attribution, whereas they should re-invest in the last 2 suggested campaigns as these are the ones responsible to get more visitors to effectively place a purchase order as a last-touch-attribution.

5 campaigns to re-invest	Why this campaign?
1) Interview-cool-tshirts-founder	1st to attract more initial visitors
2) Getting-to-know-cool-tshirts	2nd to attract more initial visitors
3) Ten-crazy-cool-tshirts-facts	3rd to attract more initial visitors
4) Weekly-newsletter	1st to get visitors to make a purchase
5) Retargeting-ad	2nd to get visitors to make a purchase

Thanks!



Learning SQL from Scratch
CAPSTONE – Attribution Queries
Mafalda Patrício
12/11/18