

LEARN SQL FROM SCRATCH

Nov 27, 2018 – Jan 30, 2019

LUKASZ BOGACZYK

First- And Last-Touch Attribution Project

1. Get familiar with the company.

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?

Get familiar with CoolTShirts

query typed against database

```

1  SELECT COUNT (DISTINCT utm_campaign) AS
   'number_of_campaigns'
2  FROM page_visits
3  ;
4
5  SELECT COUNT (DISTINCT utm_source) AS
   'number_of_sources'
6  FROM page_visits
7  ;
8
9  SELECT DISTINCT utm_source,
10 utm_campaign
11 FROM page_visits
12 ORDER BY utm_source
13 ;
14

```

query results

number_of_campaigns	
8	
number_of_sources	
6	
utm_source	utm_campaign
buzzfeed	ten-crazy-cool-tshirts-facts
email	weekly-newsletter
email	retargetting-campaign
facebook	retargetting-ad
google	paid-search
google	cool-tshirts-search
medium	interview-with-cool-tshirts-founder
nytimes	getting-to-know-cool-tshirts
nytimes	getting-to-know-cool-tshirts
medium	interview-with-cool-tshirts-founder
google	cool-tshirts-search
google	weekly-newsletter

Get familiar with CoolTShirts

CoolTShirts uses 8 different campaigns and 6 different media (sources) to get to customers. There are different campaigns that use one source to reach to customers – weekly-newsletter and retargeting-campaign are released via email, and paid-search together with cool-thirts-search are related to google source.

```
1 SELECT DISTINCT utm_source,  
2 utm_campaign  
3 FROM page_visits  
4 WHERE utm_source IN ('email', 'google')  
5 ;  
6
```

utm_source	utm_campaign
email	weekly-newsletter
email	retargeting-campaign
google	paid-search
google	cool-tshirts-search
google	cool-tshirts-search

Campaign is like a theme, a set of various activities related to each other in order to get particular feedback from a customer (in our case making people to get to the website and buy one of our fancy and cool t-shirts.)

On the other hand source is the mean of conveying the ideas linked to a campaign. In our case we use Internet technologies, i.e. newsletters, paid searches, articles published across various sites.

Get familiar with CoolTShirts

```
1 SELECT DISTINCT page_name
2 FROM page_visits
3 ;
```

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

To check some other queries against the database, and gather more info, it may come in handy to know how many distinct clicks we gained on each page (to later calculate conversion rates).

```
1 SELECT COUNT (DISTINCT user_id) AS
   'dist_visits_per_page',
2 page_name
3 FROM page_visits
4 GROUP BY 2
5 ORDER BY 1 DESC
6 ;
```

dist_visits_per_page	page_name
1979	1 - landing_page
1881	2 - shopping_cart
1431	3 - checkout
361	4 - purchase

Get familiar with CoolTShirts

What I found interesting is that some users have more than 4 actions on our site, so they visited `landing_page` and `shopping_cart` more than once before they decided to continue to checkout and make a purchase.

One may spot different numbers of total visits per `landing_page` and `shopping_cart` when compared to unique ones.



```

1 SELECT COUNT (*) AS 'total_visits'
2 FROM page_visits
3 GROUP BY page_name
4 ;
5 SELECT COUNT (DISTINCT user_id) AS 'unique_visits'
6 FROM page_visits
7 GROUP BY page_name
8 ;

```

total_visits
2000
1900
1431
361
unique_visits
1979
1881
1431
361

What is the user journey?

How many first touches is each campaign responsible for?

When we try to get first-touch-visits per campaign we see that only 4 out of 8 campaigns generated the clicks. On the other hand all campaigns are listed when it comes to checking unique clicks per purchase page. For further analysis one of the question may be:

- Were some of those campaigns designed just to persuade user to buy a t-shirt when he is already familiar with our products? Or maybe they are simply unefficient?

ft_campaign	name_of_campaign	source
622	interview-with-cool-tshirts-founder	medium
612	getting-to-know-cool-tshirts	nytimes
576	ten-crazy-cool-tshirts-facts	buzzfeed
169	cool-tshirts-search	google
169	cool-tshirts-search	google

```

1 WITH first_touch AS (
2     SELECT user_id,
3         MIN (timestamp) AS 'first_touch_at'
4     FROM page_visits
5     GROUP BY user_id)
6 SELECT COUNT (ft.user_id) AS 'ft_per_campaign',
7     pv.utm_campaign AS 'name_of_campaign'
8 FROM first_touch ft
9 JOIN page_visits pv
10     ON ft.user_id = pv.user_id
11     WHERE ft.first_touch_at = pv.timestamp
12 GROUP BY 2
13 ORDER BY 1 DESC
14 ;

```

```

15 ;
16 ORDER BY 1 DESC
17 GROUP BY 2
18 WHERE first_touch_at = timestamp

```

What is the user journey?

How many first touches is each campaign responsible for?

Now I would like to present how many clicks we have per landing_page and purchase per each campaign, so to clarify that some sources get more credit when we take into consideration last-touch that ends with moving to the purchase page.

```

1  SELECT COUNT (*) AS 'total_visits',
2     utm_campaign,
3     utm_source
4  FROM page_visits
5  WHERE page_name = '1 - landing_page'
6  GROUP BY 2
7  ORDER BY 1 DESC
8  ;
9
10 SELECT COUNT (*) AS 'total_visits',
11     utm_campaign,
12     utm_source
13 FROM page_visits
14 WHERE page_name = '4 - purchase'
15 GROUP BY 2
16 ORDER BY 1 DESC
17 ;

```

total_visits	utm_campaign	utm_source
625	interview-with-cool-tshirts-founder	medium
617	getting-to-know-cool-tshirts	nytimes
587	ten-crazy-cool-tshirts-facts	buzzfeed
171	cool-tshirts-search	google
total_visits	utm_campaign	utm_source
115	weekly-newsletter	email
113	retargeting-ad	facebook
54	retargeting-campaign	email
52	paid-search	google
9	getting-to-know-cool-tshirts	nytimes
9	ten-crazy-cool-tshirts-facts	buzzfeed
7	interview-with-cool-tshirts-founder	medium
2	cool-tshirts-search	google
5	cool-tshirts-search	google
1	interview-with-cool-tshirts-founder	medium
0	ten-crazy-cool-tshirts-facts	buzzfeed
3	getting-to-know-cool-tshirts	nytimes

What is the user journey?

How many first touches is each campaign responsible for?

Campaigns released via email and facebook source are responsible for the movement at checkout and purchase site. As far as I am concerned, a user who gets a newsletter or is influenced by retargeting campaign has already made a subscription to our site (or Google collected the data from our cookies and now shows each of customers attracted by our site some paid-ads related to our site). That is why those four campaigns may have different aims.

Therefore those campaigns may seem not quite efficient when it comes to get new customers, but they may be crucial in convincing people to check our site once again and buy some products.

Efficient campaigns to attract users (interview at medium.com, an article at NY Times, some facts on BuzzFeed) appear to have almost none effect to generate last-touch visits.

This may be a good point to think of any future campaign theme.

As the interview and some facts about the company aroused clients attention, maybe we can think of a virtual walk through the office? So clients can see how we work, deal with orders, how the tricks of the trade look like.

What is the user journey?

How many first touches is each campaign responsible for?

On the other hand following campaigns: weekly-newsletter, retargetting-campaign (both released via mail source), retargetting-ad campaign and paid-search give satisfactory results to generate last-touch visits.

In terms of investing money in future campaigns we should take into consideration that division as well as the aim of particular campaign.

What is the user journey?

Attribution model diversity

I searched the Internet to find any articles related to first-and-touch visits, to get some more info how those attributes are used by companies, to check if there are other approaches when forming attribution analysis. What is more, to get the answer for my question from the previous slide – could campaigns be designed to generate particular clicks?

I checked how Google deals with that. In their [attribution model example](#) one may spot that any single click may be associated to a diverse attribution model. Furthermore, one may notice approaches linked with linear analysis or time decay model.

What is the user journey?

Attribution model diversity

Google Attribution Modeling Overview



Last Interaction attribution model



Last Non-Direct Click



Last Google Ads Click – they count the first and only click related to paid search channel (we also use same channel via google engine, but we do not take into account such complex approach)



First Interaction



Linear Model

What is the user journey?

Attribution model diversity

Google Attribution Modeling Overview



Time Decay – this is the one that fits my conclusion about email campaigns and facebook retargeting-ad as it is explained as – *model that gives more credit to the touchpoints closest to the conversion. It makes the assumption that the closer to the conversion, the more influence it had on the conversion.*

Results of queries confirm my initial thoughts about type of campaign and source being used – email and facebook gets more credit as it generated more clicks on the checkout and purchase page. That is why we should not discount them for being ineffective for first-touch visits as their purpose is to be efficient to get last-touch visits.



Position Based

What is the user journey?

Attribution model diversity – thoughts

Together with some useful information from [How To Choose A Proven Marketing Attribution Model](#) I may make a conclusion that choosing an attribution model depends on the nature of our business, the data one would like to get from analysis, the complexity of campaign and sources/media used.

It is worth to mention that various attribution approaches may suit better depending on the type of campaign we run and type of source/media chosen to convey the ideas.

Instead of thinking about first-and-last touch attribution it might pay off to think of a multi-touch attribution model when we would like invest more money onto campaigns.

What is the user journey?

How many last touches is each campaign responsible for?

```

1 WITH last_touch AS (
2   SELECT utm_campaign,
3     MAX (timestamp) AS 'lt_touch_at',
4     user_id
5 FROM page_visits
6 GROUP BY user_id)
7 SELECT COUNT (lt.lt_touch_at) AS 'lt_campaign',
8   pv.utm_campaign AS 'name_of_campaign'
9 FROM last_touch lt
10 JOIN page_visits pv
11   ON lt.user_id = pv.user_id
12   WHERE lt.lt_touch_at = pv.timestamp
13 GROUP BY 2
14 ORDER BY 1 DESC
15 ;

```

lt_campaign	name_of_campaign
447	weekly-newsletter
443	retargeting-ad
245	retargeting-campaign
232	getting-to-know-cool-tshirts
190	ten-crazy-cool-tshirts-facts
184	interview-with-cool-tshirts-founder
178	paid-search
60	cool-tshirts-search

60	cool-tshirts-search
178	paid-search
184	interview-with-cool-tshirts-founder

What is the user journey?

How many visitors make a purchase?

The query from previous slide does not include landing on the purchase site, as last-touch at that point may equal visiting i.e. checkout page when the last timestamp occurred. That is why we need to narrow our results just to the purchase page.

lt_campaign	name_of_campaign
114	weekly-newsletter
112	retargeting-ad
53	retargeting-campaign
52	paid-search
9	getting-to-know-cool-tshirts
9	ten-crazy-cool-tshirts-facts
7	interview-with-cool-tshirts-founder
2	cool-tshirts-search
5	cool-tshirts-search
1	interview-with-cool-tshirts-founder
1	ten-crazy-cool-tshirts-facts

```

1  WITH last_touch AS (
2    SELECT utm_campaign,
3    MAX (timestamp) AS 'lt_touch_at',
4    user_id
5  FROM page_visits
6  GROUP BY user_id)
7  SELECT COUNT (lt.lt_touch_at) AS 'lt_campaign',
8    pv.utm_campaign AS 'name_of_campaign'
9  FROM last_touch lt
10 JOIN page_visits pv
11   ON lt.user_id = pv.user_id
12   WHERE lt.lt_touch_at = pv.timestamp
13   AND pv.page_name = '4 - purchase'
14 GROUP BY 2
15 ORDER BY 1 DESC
16 ;

```

```

17 :
18 :
19 ORDER BY 1 DESC
20 ORDER BY 1 DESC

```


What is the user journey?

How many visitors make a purchase?

To be honest with You at that point I got a bit confused.

I counted visits per purchase page (361), but when I added this line

`AND pv.page_name = '4 - purchase'` to the query with last-touch-visits I got different number – 358. It was quite hard for me to get any useful piece of advice among our Slack community, but I guess this was because of my lack of ability to explain the issue. Were have those 3 customers gone?

When I compare tables of results it is easier to spot that discrepancy. But after watching the walkthrough video I understand this is a joint result of `MAX(timestamp)` clause together with counting only those who get to the purchase page.

total_visits	lt_campaign	name_of_campaign
115	114	weekly-newsletter
113	112	retargeting-ad
54	53	retargeting-campaign
52	52	paid-search
9	9	getting-to-know-cool-tshirts
9	9	ten-crazy-cool-tshirts-facts
7	7	interview-with-cool-tshirts-founder
2	2	cool-tshirts-search
5	5	cool-tshirts-search
1	1	interview-with-cool-tshirts-founder

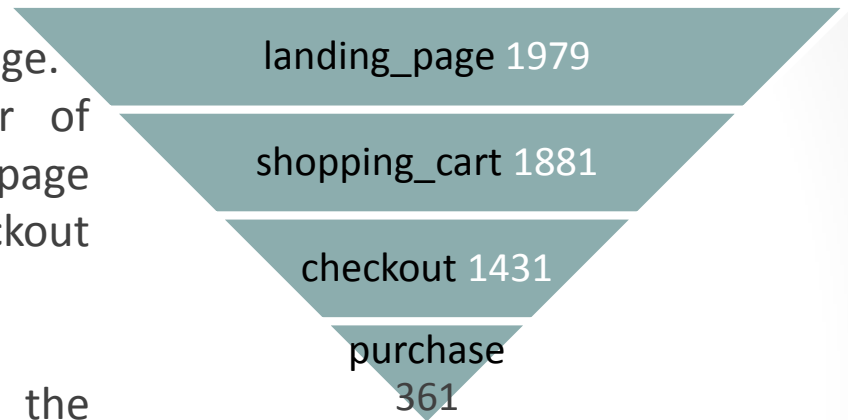
Optimize the campaign budget

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

We got 1979 unique visits per landing_page. The pyramid shows that the number of customers who get to the purchase page drops significantly, mostly between checkout and purchase page.

Issues for further consideration:

- What stops people from moving to the purchase?
- Is it related to payment and shipment options they may choose from? (high costs of products, insufficient payment forms, limited possibilities for international shipment, etc. – as those are known for customer before moving to finalize the transaction.



Optimize the campaign budget

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

As I mentioned before in order to re-invest in campaigns we should not take into account a campaign itself but we ought to consider the aim it may have together with source we would like to use. We should think complex as email and facebook campaigns failed to get first-touches, but they proved to be efficient in making people checking the site again.

That is why we should make a division of campaigns:

- Campaigns to attract new customers
- Campaigns to make people visit our site again and make a purchase

Conversion rates
per first type of
campaign (total of
1979 unique visits ➡)

ft_per_campaign		name_of_campaign
622	31,33%	interview-with-cool-tshirts-founder
612	30,92%	getting-to-know-cool-tshirts
576	29,11%	ten-crazy-cool-tshirts-facts
169	8,54%	cool-tshirts-search

Optimize the campaign budget

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

Now let's consider last-touches per campaign, but just stick to those that were successful in getting people's attention to our products.

As we get 358 users with last-touches on purchase page it is clear that the conversion rate for successful first-touch campaigns considerably decreased.

lt_campaign		name_of_campaign
114		weekly-newsletter
112		retargetting-ad
53		retargetting-campaign
52		paid-search
9	2,51%	getting-to-know-cool-tshirts
9	2,51%	ten-crazy-cool-tshirts-facts
7	1,96%	interview-with-cool-tshirts-founder
2	0,56%	cool-tshirts-search



Conversion rates
per first type of
campaign (total of
358 unique visits on
purchase page)

Optimize the campaign budget

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

However, when we count the conversion rates for campaigns of the second type (campaigns to encourage to check the site again) the numbers clearly show their effectiveness.

I would recommend to re-invest in following campaigns.

In order to arouse people's attention:

- (1) Any interview and (2) tricks-of-the-trade-like campaign with the use of medium and NY times.
- (3) Another crazy article concerning our crazy and cool t-shirts (launched via BuzzFeed)

To encourage people to check the site again

- 4) Weekly-newsletter or (5) retargetting-campaign via mail and/or facebook

lt_campaign		name_of_campaign
114	31,83%	weekly-newsletter
112	31,28%	retargetting-ad
53	14,80%	retargetting-campaign
52	14,53%	paid-search
9		getting-to-know-cool-tshirts
9		ten-crazy-cool-tshirts-facts
7		interview-with-cool-tshirts-founder
2		cool-tshirts-search



Conversion rates per second type of campaign (total of 358 unique visits on purchase page)