Learn SQL from Scratch:
First- and Last-Touch Attribution

Jerry Bohun | Jun 5, 2018 - Jul 31, 2018

CoolTShirts

CoolTShirts an innovative apparel shop, is running a bunch of marketing campaigns. Answer these questions about their campaigns:

➔ **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_campaignand utm_source.
- ◆ What pages are on their website?

➔ **2. Get familiar with the company**
- ◆ 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?

1. Get familiar with CoolTShirts

1. **How many campaigns and sources does CoolTShirts use? Which source is used for each campaign?**

❏ CoolTShirts had a total of **8 campaigns** and used **6 different sources**.
❏ Table below lists sources used for each campaign. Each campaign is assigned with only one source.

2. **What pages are on the CoolTShirts website?**

❏ There are **4 pages** on the CoolTShirts website and they are listed below.

```
1   -- Task 1a
2   SELECT COUNT(DISTINCT(utm_campaign)) as 'Total number of campaigns'
3   FROM page_visits;
4   SELECT DISTINCT(utm_campaign) as 'Campaigns'
5   FROM page_visits;
6
7   -- Task 1b
8   SELECT COUNT(DISTINCT(utm_source)) as 'Total number of sources'
9   FROM page_visits;
10  SELECT DISTINCT(utm_source) as 'Sources'
11  FROM page_visits;
12
13  --Task 1c
14  SELECT DISTINCT utm_campaign, utm_source
15  FROM page_visits;
```

Total number of campaigns
8
Campaigns
getting-to-know-cool-tshirts
weekly-newsletter
ten-crazy-cool-tshirts-facts
retargetting-campaign
retargetting-ad
interview-with-cool-tshirts-founder
paid-search
cool-tshirts-search

Total number of sources
6
Sources
nytimes
email
buzzfeed
facebook
medium
google

```
1   -- Task 2a
2   SELECT DISTINCT(page_name)
3   FROM page_visits;
```

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

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

2. What is the user journey?

3. How many first touches is each campaign responsible for?

❏ There are only 4 out of 8 campaigns that are responsible for first touch.
❏ Table below lists these campaigns and shows a total of "first touches" for each.

❏ Looking at the Hint, I don't think there was a need to show the source and include JOIN in order to get the necessary results.

```sql
1   -- Task 3
2   WITH ft AS (
3   -- List all IDs with its lowest timestamp aka first touch
4   SELECT user_id,
5          MIN(timestamp),
6          utm_campaign
7      FROM page_visits
8   -- Without GROUP BY user_id we would get only one record with lowest/earliest timestamp
9      GROUP BY user_id
10          )
11  -- NO need for a JOIN as data is available in the original table
12  SELECT  ft.utm_campaign AS 'Campaign',
13          COUNT(*) AS 'Count'
14  FROM ft
15  GROUP BY 1
16  ORDER BY 2 DESC;
```

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

2. What is the user journey?

4. How many last touches is each campaign responsible for?

❏ All 8 campaigns are responsible for last touch and table below lists the count.
❏ However, campaigns responsible for "first touches" are not the highest scorers for "last touches".

❏ Looking at the Hint, again I don't think there was a need to show the source and include JOIN in order to get the necessary results.

```
1   -- Task 4
2   WITH lt AS (
3   -- List all IDs with its highest timestamp aka last touch
4   SELECT user_id,
5           MAX(timestamp),
6           utm_campaign
7       FROM page_visits
8   -- Without GROUP BY user_id we would get only one record with highest/latest timestamp
9       GROUP BY user_id
10              )
11  -- NO need for a JOIN as data is available in the original table
12  SELECT  lt.utm_campaign AS 'Campaign',
13          COUNT(*) AS 'Count'
14  FROM lt
15  GROUP BY 1
16  ORDER BY 2 DESC;
```

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

2. What is the user journey?

5. How many visitors make a purchase?

❏ A total of 361 users hit the final purchase page.

```
1   -- Task 5
2   SELECT COUNT(DISTINCT(user_id)) AS 'Number of customers that finalised a deal'
3   FROM page_visits
4   WHERE page_name = '4 - purchase';
```

Number of customers that finalised a deal
361

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

❏ Table below lists all campaigns that are responsible for hitting the purchase page on users' last touch.
❏ It is worth noting that paid-search campaign has significantly change its position comparing to table from Task 4.

```
1   -- Task 6
2   WITH lt AS (
3   -- List all IDs with its highest timestamp aka last touch
4   SELECT user_id,
5           MAX(timestamp),
6           utm_campaign
7       FROM page_visits
8   -- Combining Task 4 and 5 we can see figures for campaigns that led to final purchase page
9       WHERE page_name = '4 - purchase'
10      GROUP BY user_id
11          )
12  -- NO need for a JOIN as data is available in the original table
13  SELECT lt.utm_campaign AS 'Campaign that led to purchase page',
            COUNT(*) AS 'Count'
    FROM lt
    GROUP BY 1
    ORDER BY 2 DESC;
```

Campaign that led to purchase page	Count
weekly-newsletter	115
retargetting-ad	113
retargetting-campaign	54
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

2. What is the TYPICAL user journey?

Typical user journey can be presented as a four stage process. Starting with a **landing page**, through a **shopping cart** to **checkout** and final **purchase** step.
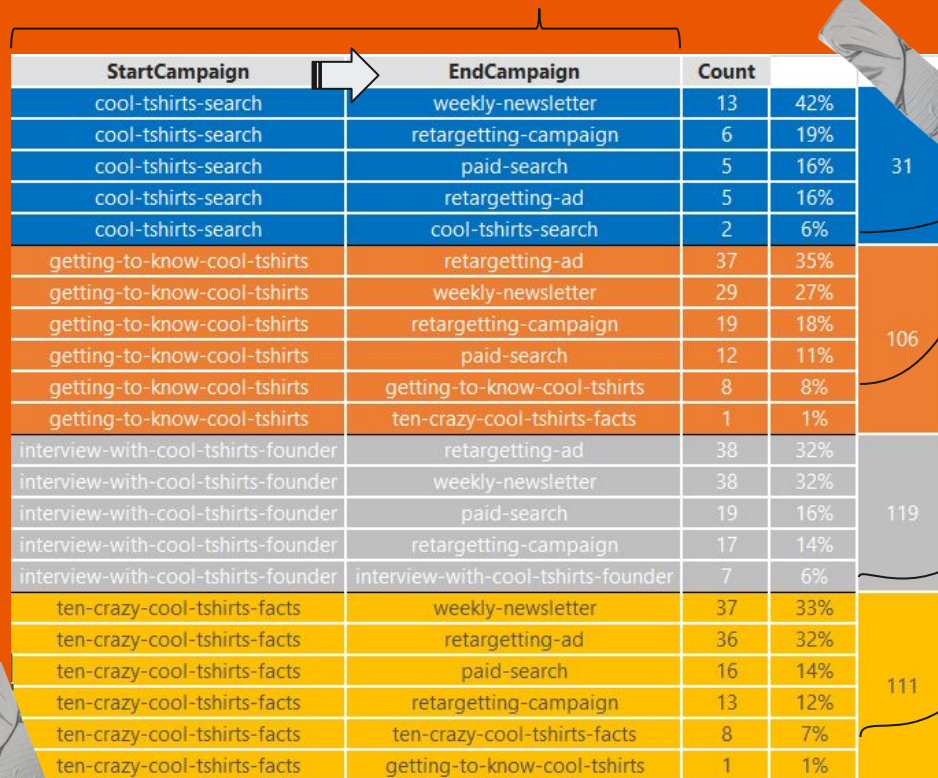
From the table below we can see that every user journey starts with four campaigns marked in yellow. Unfortunately, user engagement is dropping rapidly after second stage where user is presented with a shopping basket and only <u>1% of total hits are actually finalised at the last page</u>.

Campaign	Source	1 - landing_page		2 - shopping_cart		3 - checkout		4 - purchase
		Total hits	% of last touch	Total hits	% of last touch	Total hits	% of last touch	Total hits
interview-with-cool-tshirts-founder	medium	625	3.36%	515	25.63%	31	77.42%	7
getting-to-know-cool-tshirts	nytimes	617	6.16%	682	22.43%	41	78.05%	9
ten-crazy-cool-tshirts-facts	buzzfeed	587	5.45%	570	22.11%	32	71.88%	9
cool-tshirts-search	google	171	4.09%	133	34.59%	7	71.43%	2
paid-search	google	0		0		179	70.39%	52
retargetting-campaign	email	0		0		246	78.05%	54
retargetting-ad	facebook	0		0		445	74.38%	113
weekly-newsletter	email	0		0		450	74.00%	115

Second run of four new campaigns, marked with green above, allowed to finalise <u>additional 17%</u> of total hits to a landing page from the "starting campaigns". It's clearly rebuilding users' engagement by encouraging them to move to a third stage of the journey before placing an order.

2. What is the TYPICAL user journey?

The same user journey can be mapped from **four starting campaigns** through to the ending campaign.

StartCampaign	→	EndCampaign	Count		
cool-tshirts-search		weekly-newsletter	13	42%	
cool-tshirts-search		retargetting-campaign	6	19%	
cool-tshirts-search		paid-search	5	16%	31
cool-tshirts-search		retargetting-ad	5	16%	
cool-tshirts-search		cool-tshirts-search	2	6%	
getting-to-know-cool-tshirts		retargetting-ad	37	35%	
getting-to-know-cool-tshirts		weekly-newsletter	29	27%	
getting-to-know-cool-tshirts		retargetting-campaign	19	18%	106
getting-to-know-cool-tshirts		paid-search	12	11%	
getting-to-know-cool-tshirts		getting-to-know-cool-tshirts	8	8%	
getting-to-know-cool-tshirts		ten-crazy-cool-tshirts-facts	1	1%	
interview-with-cool-tshirts-founder		retargetting-ad	38	32%	
interview-with-cool-tshirts-founder		weekly-newsletter	38	32%	
interview-with-cool-tshirts-founder		paid-search	19	16%	119
interview-with-cool-tshirts-founder		retargetting-campaign	17	14%	
interview-with-cool-tshirts-founder		interview-with-cool-tshirts-founder	7	6%	
ten-crazy-cool-tshirts-facts		weekly-newsletter	37	33%	
ten-crazy-cool-tshirts-facts		retargetting-ad	36	32%	
ten-crazy-cool-tshirts-facts		paid-search	16	14%	111
ten-crazy-cool-tshirts-facts		retargetting-campaign	13	12%	
ten-crazy-cool-tshirts-facts		ten-crazy-cool-tshirts-facts	8	7%	
ten-crazy-cool-tshirts-facts		getting-to-know-cool-tshirts	1	1%	

❏ We can see that each starting campaign managed to close only 6% to 8% of total deals within the same campaign.

❏ Four ending campaigns managed to re-engage users and are responsible with 93% of total deals closed on the last purchase page.

2. What is the TYPICAL user journey?
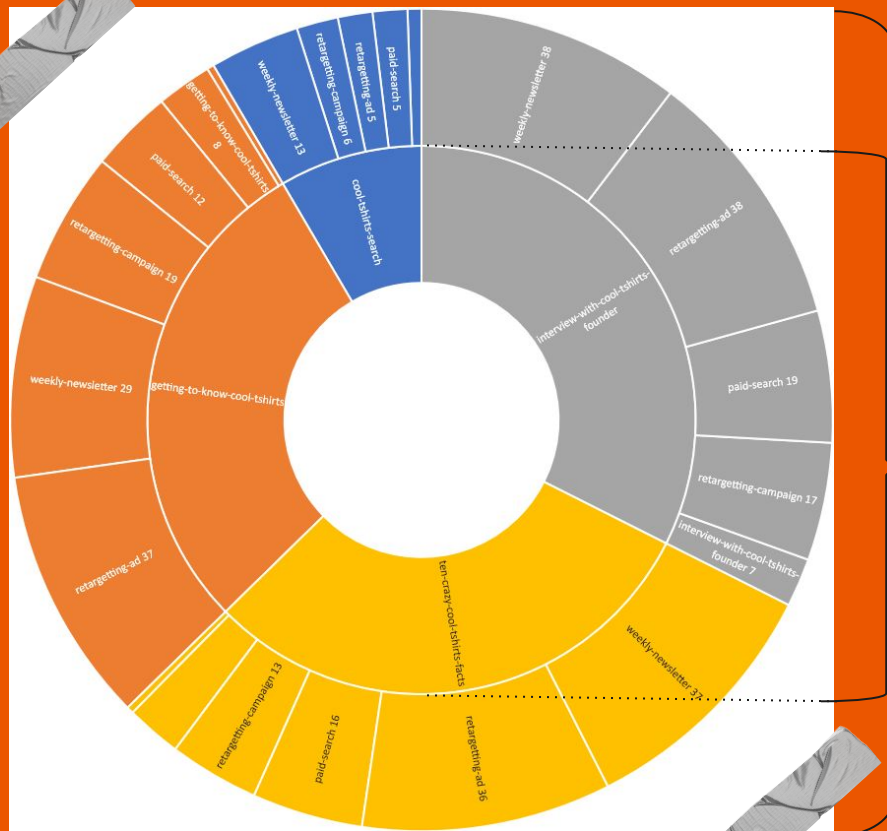
Outer circle of this data chart visualises hits on the purchase page linked to each starting campaign and divided between campaigns responsible for final step in the process

❏ A total of **361** users hit the final purchase page. However, there are **367** links between the landing and purchase page. This means that user who finished the journey in one of 8 campaigns could have started it in more than one out of 4 starting campaigns.

Inner circle of this data chart represents hits on the landing page divided between four starting campaigns.

❏ There is a total of **2000** hits on the landing page. However, there are **1979** first touches shared between 4 starting campaigns. This means that users could start the journey more than once.

Three campaigns are responsible for **91%** of total traffic to the landing page. The same three campaigns led to **92%** of total closed transactions on final purchase page through itself or one of 4 closing campaigns.

3. Optimize the campaign budget

It is very hard to advise on budget without knowing the actual financial figures.
We don't know the cost of each of the campaign or the effort and resources it consumed.
As an example we can notice below that "cool-tshirts-search" and "paid- search" campaign used google as a source. Google was the only source to bring potential users attention and finalise a deal at the final purchase page. Maybe a joined campaign should be further explored?
"Paid-search" has the lowest percentage of the last touch on checkout page, so there are higher chances to close a transaction in the final purchase page.

Campaign	Source	1 - landing_page		2 - shopping_cart		3 - checkout		4 - purchase
		Total hits	% of last touch	Total hits	% of last touch	Total hits	% of last touch	Total hits
interview-with-cool-tshirts-founder	medium	625	3.36%	515	25.63%	31	77.42%	7
getting-to-know-cool-tshirts	nytimes	617	6.16%	682	22.43%	41	78.05%	9
ten-crazy-cool-tshirts-facts	buzzfeed	587	5.45%	570	22.11%	32	71.88%	9
cool-tshirts-search	google	171	4.09%	133	34.59%	7	71.43%	2
paid-search	google	0		0		179	70.39%	52
retargetting-campaign	email	0		0		246	78.05%	54
retargetting-ad	facebook	0		0		445	74.38%	113
weekly-newsletter	email	0		0		450	74.00%	115

3. Optimize the campaign budget
Starting campaigns

However, using data available to us we can clearly highlight top performing campaigns. *Getting-to-know-cool-tshirts*, *interview-with-cool-tshirts-founder* and *ten-crazy-cool-tshirts-facts* are without a question best in attracting users to our landing page and adding product to a shopping cart.These campaigns are responsible for 91% of total traffic to the landing page. The same three led to 92% of total closed transactions on final purchase page through itself or one of 4 closing campaigns. We definitely need this starting figures to be high.

Campaign	Source	1 - landing_page		2 - shopping_cart		3 - checkout		4 - purchase
		Total hits	% of last touch	Total hits	% of last touch	Total hits	% of last touch	Total hits
interview-with-cool-tshirts-founder	medium	625	3.36%	515	25.63%	31	77.42%	7
getting-to-know-cool-tshirts	nytimes	617	6.16%	682	22.43%	41	78.05%	9
ten-crazy-cool-tshirts-facts	buzzfeed	587	5.45%	570	22.11%	32	71.88%	9
cool-tshirts-search	google	171	4.09%	133	34.59%	7	71.43%	2
paid-search	google	0		0		179	70.39%	52
retargetting-campaign	email	0		0		246	78.05%	54
retargetting-ad	facebook	0		0		445	74.38%	113
weekly-newsletter	email	0		0		450	74.00%	115

3. Optimize the campaign budget

Ending campaigns

We are now left with three starting campaigns and four ending campaigns. Looking at the figures below *retargetting-ad* and *weekly-newsletter* are together closing over 60% for all transactions. In this example numbers don't lie and different choice can be risky.

StartCampaign	EndCampaign	Count	
getting-to-know-cool-tshirts	retargetting-ad	37	35%
getting-to-know-cool-tshirts	weekly-newsletter	29	27%
getting-to-know-cool-tshirts	retargetting-campaign	19	18%
getting-to-know-cool-tshirts	paid-search	12	11%
getting-to-know-cool-tshirts	getting-to-know-cool-tshirts	8	8%
getting-to-know-cool-tshirts	ten-crazy-cool-tshirts-facts	1	1%
interview-with-cool-tshirts-founder	retargetting-ad	38	32%
interview-with-cool-tshirts-founder	weekly-newsletter	38	32%
interview-with-cool-tshirts-founder	paid-search	19	16%
interview-with-cool-tshirts-founder	retargetting-campaign	17	14%
interview-with-cool-tshirts-founder	interview-with-cool-tshirts-founder	7	6%
ten-crazy-cool-tshirts-facts	weekly-newsletter	37	33%
ten-crazy-cool-tshirts-facts	retargetting-ad	36	32%
ten-crazy-cool-tshirts-facts	paid-search	16	14%
ten-crazy-cool-tshirts-facts	retargetting-campaign	13	12%
ten-crazy-cool-tshirts-facts	ten-crazy-cool-tshirts-facts	8	7%
ten-crazy-cool-tshirts-facts	getting-to-know-cool-tshirts	1	1%

3. Optimize the campaign budget

Final FIVE campaigns

With three starting campaigns and two ending campaigns we have hopefully made a significant savings in the budget and our results are still sustainable. We have lost about 9% of the traffic to our landing page.

Campaign	Source	1 - landing_page		2 - shopping_cart		3 - checkout		4 - purchase
		Total hits	% of last touch	Total hits	% of last touch	Total hits	% of last touch	Total hits
interview-with-cool-tshirts-founder	medium	625	3.36%	515	25.63%	31	77.42%	7
getting-to-know-cool-tshirts	nytimes	617	6.16%	682	22.43%	41	78.05%	9
ten-crazy-cool-tshirts-facts	buzzfeed	587	5.45%	570	22.11%	32	71.88%	9

However, with only two ending campaigns we are able to finalise only 65% of the originally closed deals.

StartCampaign	EndCampaign	Count		
getting-to-know-cool-tshirts	retargetting-ad	37	49%	
getting-to-know-cool-tshirts	weekly-newsletter	29	39%	
getting-to-know-cool-tshirts	getting-to-know-cool-tshirts	8	11%	75
getting-to-know-cool-tshirts	ten-crazy-cool-tshirts-facts	1	1%	
interview-with-cool-tshirts-founder	retargetting-ad	38	46%	
interview-with-cool-tshirts-founder	weekly-newsletter	38	46%	83
interview-with-cool-tshirts-founder	interview-with-cool-tshirts-founder	7	8%	
ten-crazy-cool-tshirts-facts	weekly-newsletter	37	45%	
ten-crazy-cool-tshirts-facts	retargetting-ad	36	44%	82
ten-crazy-cool-tshirts-facts	ten-crazy-cool-tshirts-facts	8	10%	
ten-crazy-cool-tshirts-facts	getting-to-know-cool-tshirts	1	1%	