code|cademy

Capstone Attribution Queries

Learn SQL from Scratch
Chris Beyer
May 31, 2019

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 Get familiar with CoolTShirts

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.
- To solve this question we will need to write 3 queries; see queries below
- There are 8 campaigns and 6 sources
- UTM parameters are a way of tracking visits to a website, If an ad campaign drives a lot of visits to their site, then they know that source is working! We say that those visits are attributed to the ad campaign. CoolTShirts site owner used special links containing UTM parameters in their campaigns and sources. When a user clicks one, a row is added to a database describing their page visit. Such as the weekly-newsletter campaign is sourced from an email and cool-tshirts-search is sourced from google.

```
/*
Distinct campaigns
*/
SELECT COUNT (DISTINCT utm_campaign)
FROM page_visits;
/*
Distinct sources
*/
SELECT COUNT (DISTINCT utm_source)
FROM page_visits;
/*
how they are related
*/
SELECT DISTINCT utm_campaign, utm_source
FROM page_visits;
```

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

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 Get familiar with CoolTShirts

What pages are on their website?
- Landing Page
- Shopping Cart
- Checkout
- Purchase

```
/*
distinct values of the page_name column
*/
SELECT DISTINCT page_name
FROM page_visits;
```

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

2. What is the user journey?

2.1 What is the user journey?

How many first touches is each campaign responsible for?

user_id	first_touch_at	utm_source	utm_campaign	COUNT (utm_campaign)
99684	2018-01-13 13:20:49	google	cool-tshirts-search	169
99765	2018-01-04 05:59:46	buzzfeed	ten-crazy-cool-tshirts-facts	576
99933	2018-01-25 00:04:39	nytimes	getting-to-know-cool-tshirts	612
99990	2018-01-13 23:30:09	medium	interview-with-cool-tshirts-founder	622

```
/*
How many First touches is each campaign responsible for
query
*/
WITH first_touch AS (
    SELECT user_id,
        MIN(timestamp) as first_touch_at
    FROM page_visits
    GROUP BY user_id)
SELECT ft.user_id,
    ft.first_touch_at,
    pv.utm_source,
    pv.utm_campaign,
    COUNT (utm_campaign)
FROM first_touch ft
JOIN page_visits pv
    ON ft.user_id = pv.user_id
    AND ft.first_touch_at = pv.timestamp
GROUP BY utm_campaign
ORDER BY 5 ASC;
```

2.2 What is the user journey?

How many last touches is each campaign responsible for?

user_id	last_touch_at	utm_source	utm_campaign	COUNT (utm_campaign)
99344	2018-01-18 21:36:32	google	cool-tshirts-search	60
98840	2018-01-10 04:58:48	google	paid-search	178
99838	2018-01-02 07:40:34	medium	interview-with-cool-tshirts-founder	184
99765	2018-01-04 05:59:47	buzzfeed	ten-crazy-cool-tshirts-facts	190
99589	2018-01-15 04:55:43	nytimes	getting-to-know-cool-tshirts	232
99990	2018-01-16 11:35:09	email	retargetting-campaign	245
99928	2018-01-24 05:26:09	facebook	retargetting-ad	443
99933	2018-01-26 06:18:39	email	weekly-newsletter	447

```
/*
How many last touches is each campaign responsible for
query
*/
WITH last_touch AS (
    SELECT user_id,
        Max(timestamp) as last_touch_at
    FROM page_visits
    GROUP BY user_id)
SELECT lt.user_id,
    lt.last_touch_at,
    pv.utm_source,
    pv.utm_campaign,
    COUNT (utm_campaign)
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
GROUP BY utm_campaign
ORDER BY 5 ASC;
```

2.3 What is the user journey?

How many visitors make a purchase?

COUNT (DISTINCT user_id)
361

```
/*
How many vistors make a purchase
*/
SELECT COUNT (DISTINCT user_id)
FROM page_visits
where page_name = '4 - purchase';
```

2.4 What is the user journey?

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

user_id	last_touch_at	utm_source	utm_campaign	COUNT (utm_campaign)
95650	2018-01-18 00:25:00	google	cool-tshirts-search	2
83547	2018-01-10 18:20:21	medium	interview-with-cool-tshirts-founder	7
92172	2018-01-16 15:15:29	nytimes	getting-to-know-cool-tshirts	9
98651	2018-01-15 04:17:36	buzzfeed	ten-crazy-cool-tshirts-facts	9
94567	2018-01-19 16:37:58	google	paid-search	52
99285	2018-01-24 09:00:58	email	retargetting-campaign	54
99897	2018-01-06 09:41:19	facebook	retargetting-ad	113
99933	2018-01-26 06:18:39	email	weekly-newsletter	115

```
WITH last_touch AS (
    SELECT user_id,
        Max(timestamp) as last_touch_at
    FROM page_visits
            where page_name = '4 - purchase'
    GROUP BY user_id)
SELECT lt.user_id,
    lt.last_touch_at,
    pv.utm_source,
    pv.utm_campaign,
    COUNT (utm_campaign)
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
GROUP BY utm_campaign
ORDER BY 5 ASC;
```

2.5 What is the user journey?

What is the typical user journey?
- A typical user journey for CoolTShirts is when a user clicks on the various ads(campaigns) over 31.4% clicked on interview-with-cool-tshirts-founder campaign. However this is only part of the user experience when you look at the last touch campaigns 22.6% of users journey through the weekly-newsletter campaign which is sourced from email. This is vital information to know but what is truly important in our analysis is how many of these users from the various campaigns make a purchase in this case it is 361 of them do. But the most successful campaigns that lead to a purchase are from the weekly-newsletters 31.9% which is sourced by email and a retargeting-ad 31.3% which is sourced by Facebook. This means from over 63.2% of purchases come from these campaigns. That is a significant percentage of purchases and should tell how we should focus our budget more in these areas versus much lower performing campaigns.

3. Optimize the campaign budget

3.1 Optimize the campaign

CoolTShirts can re-invest in 5 campaigns. Which should they pick and why?
- I would recommend that CoolTShirts reinvest in the following campaigns in order of highest reinvestment:
- weekly_newsletter, retargetting-ad, retargetting-campaign, paid-search, and getting-to-know-cool-tshirts
- Why I would recommend these campaigns and this order is because the top 3 campaigns; weekly_newsletter, retargetting-ad, retargetting-campaign account for 78.1% of purchases and are the same top 3 in last touch campaigns and account for 57.3%. While paid-search has a much lower last touch campaign result yet even out of those 178 touches they convert very well in purchases accounting for 52 and 14.4% of total purchases. Last is getting-to-know-cool-tshirts because it has 232 last touches while it does the same in purchases as ten-crazy-cool-tshirts-facts it has a significantly higher chance to do better based on the higher number of last touch points so if they were to put more budget against it I believe it will have a higher conversion of purchases.