codecademy

Analyzing Attribution

Learn SQL from Scratch

Craig Heier | 2018.08.21

Table of Contents

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**
- **USER JOURNEY ANALYSIS**
- **CAMPAIGN OPTIMIZATION**

1. Get Familiar with CoolTShirts

How many campaigns and sources does CoolTShirts use and how are they related? Which source is used for each campaign?

- Campaigns are specific marketing messaging designed to promote a specific product, service or action by the targeted prospects.
- A source is a type of delivery method used to communicate the campaign message. For example, an email blast, Facebook ads or search engines like Google.
- Sources such as Google & Email run multiple campaigns in the dataset. For example, Google was used for two different campaigns: ‘paid-search’ & ‘cool-tshirts-seach’.

```sql
--1. Get Familiar with CoolTShirts
-- Get COUNT of Campaigns & Count of Sources in one query.
SELECT COUNT(DISTINCT utm_campaign) AS 'Campaign Count',
       COUNT(DISTINCT utm_source) AS 'Source Count'
FROM page_visits;
```

Campaign Count	Source Count
8	6

```sql
--List of sources per campaign. (Shows relationship)
SELECT DISTINCT utm_campaign AS Campaign,
  utm_source AS Source
FROM page_visits;
```

Campaign	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.a CoolTShirts – Website Pages

What different pages are on the CoolTShirts website?

- Page names in this dataset are a subset of the individual pages on the website. They are broken up into four types to capture a user's journey from the first visit to a landing page to their last with the purchase page.
- The query to generate this list must have the SELECT DISTINCT statement to assure the results only return unique values.

```
21  --1.a List of Pages on the CoolTShirts website
22  SELECT DISTINCT page_name AS Webpages
23  FROM page_visits;
```

Webpages
1 – landing_page
2 – shopping_cart
3 – checkout
4 – purchase

2. What is the User Journey – First Touch

How many first touches is each campaign responsible for?

- First-touch attribution only considers the first `utm_campaign` for each customer. This is how we identify which campaigns initially draw visitors to the website.
- The first-touch is identified by querying the timestamps with the MIN date for each user.
- The top three campaigns that drove users to site were all content based such as an interview with the CoolTShirts founder and informational articles about the company.

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

```sql
--2. Count how many First touches each campaign is responsible for.
---Create temp table to capture first touch by user id.
WITH first_touch AS (
  SELECT user_id,
    MIN(timestamp) AS first_touch_at
  FROM page_visits
  GROUP BY user_id),
--Create second temp table called "first_attributes" which adds
--the campaign and source attributes joins them to the first temp table
--on userid and timestamp.
first_attributes AS (
  SELECT ft.user_id,
      ft.first_touch_at,
      pv.utm_source,
      pv.utm_campaign
  FROM first_touch ft
  JOIN page_visits pv
      ON ft.user_id = pv.user_id
      AND ft.first_touch_at = pv.timestamp
)
--Count number of rows where first touch is associated
--with a campaign and source.
SELECT first_attributes.utm_campaign AS Campaign,
       COUNT(*) AS 'Count'
FROM first_attributes
GROUP BY 1
ORDER BY 2 DESC;
```

2.a What is the User Journey – Last Touch

How many last touches is each campaign responsible for?

- Last-touch attribution only considers the last `utm_campaign` for each customer. This is how we identify which campaigns draw visitors back to the website, especially for making a final purchase.
- The last-touch is identified by querying the timestamps with the MAX date for each user.

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

```sql
--2.a Count how many Last touches each campaign is responsible for.
---Create temp table to capture last touch by user id.
WITH last_touch AS (
  SELECT user_id,
    MAX(timestamp) AS last_touch_at
  FROM page_visits
  GROUP BY user_id),
--Create second temp table called "last_attributes" which adds
--the campaign and source attributes and joins them to the
--first temp table on userid and time stamp.
last_attributes AS (
  SELECT lt.user_id,
      lt.last_touch_at,
      pv.utm_source,
      pv.utm_campaign
  FROM last_touch lt
  JOIN page_visits pv
      ON lt.user_id = pv.user_id
      AND lt.last_touch_at = pv.timestamp
)
--Count number of rows where first touch is associated
--with a campaign and source.
SELECT last_attributes.utm_campaign AS Campaign,
       COUNT(*) AS 'Count'
FROM last_attributes
GROUP BY 1
ORDER BY 2 DESC;
```

2.b What is the User Journey – Purchase Page Conversion

How many visitors make a purchase?

- Out of the 1979 users that visited the landing page, only 361 users made a purchase.
- There is a **18.24%** conversion rate from visitors on the landing page becoming customers on the purchase page.

Purchasing Users	Landing Page Users	Conversion %
361	1979	18.24%

```sql
--Count how many users visited the purchase page
SELECT COUNT(DISTINCT user_id) AS 'Purchasing Users'
FROM page_visits
WHERE page_name = '4 - purchase';

--Calculate conversion rate.
--Built temp table "p" to generate count of purchasers
WITH p AS (
  -- COUNT statement is divided by 1.0 to convert result into an integer.
  SELECT COUNT(DISTINCT user_id) / 1.0 AS purchasing_user
  FROM page_visits
  WHERE page_name = '4 - purchase'
),
--Build second temp table "t" to generate count of users
--visiting the landing page
t AS (
  SELECT page_name, COUNT(DISTINCT user_id) / 1.0 AS landing
  FROM page_visits
  WHERE page_name = '1 - landing_page'

)
--Calculating the conversion rate is Purchasers divided by total count of user
--that visited the landing page. Multipy the result by 100.0 for a percentage format.
SELECT p.purchasing_user,
       t.landing,
       ROUND(((p.purchasing_user / t.landing ) * 100.0), 2) as 'Conversion %'

FROM p, t;
```

2.c What is the User Journey – Purchase Page Attribution

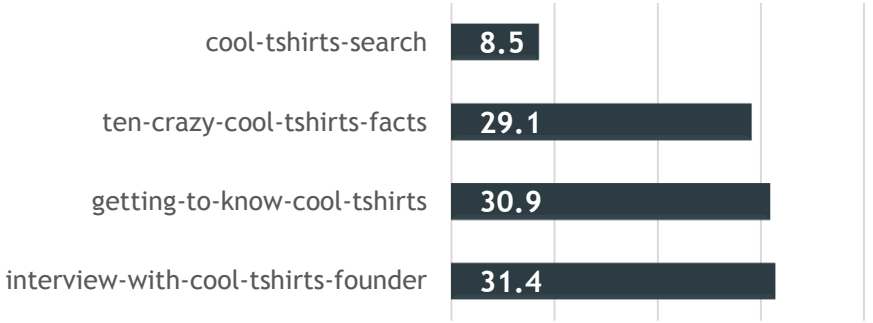
How many last touches on the purchase page is each campaign responsible for?

- For each user's session, we take the MAX timestamp where the page name is "purchase_page" and count that number by campaign.

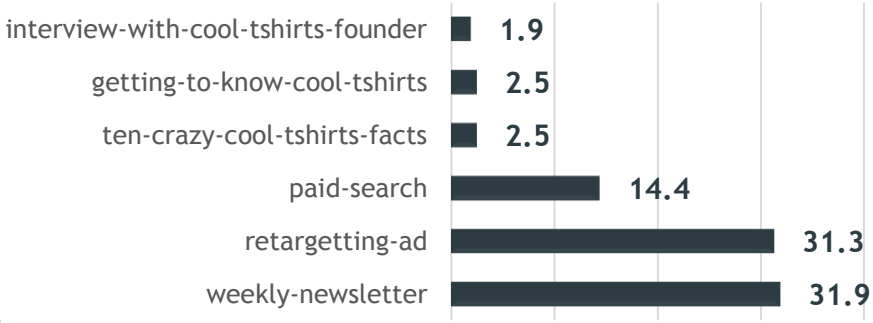
Source	Campaign	Count
email	weekly-newsletter	115
facebook	retargetting-ad	113
email	retargetting-campaign	54
google	paid-search	52
buzzfeed	ten-crazy-cool-tshirts-facts	9
nytimes	getting-to-know-cool-tshirts	9
medium	interview-with-cool-tshirts-founder	7
google	cool-tshirts-search	2

```sql
------------------------------------------------------------
--2.c Count how many last touches each campaign is responsible for on the Purchase Page.
---Create temp table to capture last touch by user id for only the Purchase Page.
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),
--Create second temp table called "last_attributes" which adds the campaign and source attributes
--and joins them to the first temp table on userid and time stamp.
last_attributes AS (
  SELECT lt.user_id,
    lt.last_touch_at,
    pv.utm_source,
    pv.utm_campaign
  FROM last_touch lt
  JOIN page_visits pv
      ON lt.user_id = pv.user_id
      AND lt.last_touch_at = pv.timestamp
)
--Count number of rows where last touch is associated with a campaign and source.
SELECT last_attributes.utm_source AS Source,
       last_attributes.utm_campaign AS Campaign,
       COUNT(*) AS 'Count'
FROM last_attributes
GROUP BY Source, Campaign
ORDER BY 3 DESC;
```

3. What is the Typical User Journey



The best performing campaigns for **First Touch attribution** are "Content" based.



The best performing campaigns for **Last Touch attribution** are "Retargetting" based.

4. Optimizing the Budget

- The top 5 campaigns that should be invested in next year include all 3 “Content” campaigns, highlighted in Yellow. The remaining two spots should go to the top campaign that drove purchases with the weekly-newsletter then the top “retargetting” campaign. (Both highlighted in Green).
- The 3 “Content” campaigns drive new visitors to the site and increase brand recognition. First impressions are worth the investment.
- In addition to “Content” campaigns, the investment of “retargetting” campaigns will endorse the full user journey.
- *Interesting to note that the “Paid Search” campaign produced absolutely zero first touches. The company website’s SEO performance might need to be evaluated if PPC ads can’t even drive page views.

Campaign	1st Touches	% - 1st Touches	Last Touches	% - Last Touches	Purchases	% - Purchases
Interview-with-cool-tshirts-founder	622	31%	184	9%	7	2%
Getting-to-know-cool-tshirts	612	31%	232	12%	9	2%
Ten-crazy-cool-tshirt-facts	576	29%	190	10%	9	2%
Cool-tshirts-search	169	9%	60	3%	2	1%
Weekly-newsletter	0	0%	447	23%	115	32%
Retargetting-ad	0	0%	443	22%	113	31%
Retargetting-campaign	0	0%	245	12%	54	15%
Paid-search	0	0%	178	9%	52	14%