codecademy

Learn SQL from Scratch

Adrian Bee

2018-06-09

Table of Contents

1. Get familiar with CoolTShirts.com

1.1 How many campaigns and sources does CoolTShirts use and how are they related?

To obtain an overview of the campaigns and sources that CoolTShirts use we can create three simple SQL queries; the first to count the number of distinct campaigns (utm.campaign), the second the number of distinct sources (utm_source) and finally to list the distinct campaigns against the sources used. By using SELECT DISTINCT we return only unique values., for example individual campaign names..

•CoolTShirts employ 8 campaigns. A campaign is the promotion of the company and its products for the purpose of customer acquisition and company positioning.

•These campaigns utilize 6 sources. These are the type of site or means of communication used to bring the campaign to the prospective customer. In this case the sources include emails, social media platforms and online news sources. Campaigns are not limited to being run across only one source, and sources can run multiple campaigns.

•The utm.campaign column in the page_visits table we are drawing data from lists the marketing promotion, and the utm_source column lists the method and/or site of communicating the campaign.

```
--What is the number of distinct campaigns?

SELECT COUNT (DISTINCT utm_campaign)
FROM page_visits;

--How many distinct sources are used?

SELECT COUNT (DISTINCT utm_source)
FROM page_visits;

--How are the campaigns and sources related?

SELECT DISTINCT utm_campaign,
        utm_source
FROM page_visits
GROUP BY utm_campaign;
```

Campaigns	8
Sources	6

Campaign	Source
cool-tshirts-search	google
getting-to-know-cool-tshirts	nytimes
interview-with-cool-tshirts-founder	medium
paid-search	google
retargetting-ad	facebook
retargetting-campaign	email
ten-crazy-cool-tshirts-facts	buzzfeed
weekly-newsletter	email

1. Get familiar with CoolTShirts.com - continued

1.2 What pages are on their website?

The individual page names can be found using a SELECT DISTINCT query to identify unique page names (page_name). Using this we find that there are four distinct page names on the CoolTShirts website; the **landing page, shopping cart, checkout**, and **purchase pages.**

```
--Query to find number and name of pages on
CoolTShirts website

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 How many first touches is each campaign responsible for?

By using the **COUNT** function to count unique **user_id,** we can identify which campaign and source led the prospective customer to first click the link to the CoolTShirts website. We can find this information by utilizing a query that identifies the **Minimum (MIN)** timestamp for each user, and then using the **GROUP BY** clause to group the results by source and campaign. Here we discover that the source '**medium**' using the campaign of an '**interview with the founder of CoolTShirts**' accounts for highest number of first touches, with **622**.

The query we use here is more complicated that our initial four and involves creating two temporary tables, using the **WITH** and **AS** clauses to give each table an **alias**. The **GROUP BY** clause is added so as to arrange the data into groups, the **ORDER BY** clause then sorts the data by a particular column, in this case by the **COUNT** in descending (DESC) order of value. The first temporary table **(first_touch**) identifies first touches by **user_id.** The second temporary table adds the source and campaign columns to the same set by joining the **page_visits** table with the **first_touch** table on **user_id** and **timestamp**. These enable a **COUNT** of the rows where there was a first touch attributed to a campaign and source.

```
--Query to find first_touch count by campaign using temporary tables.
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.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
  )
SELECT ft_attr.utm_source,
     ft_attr.utm_campaign,
     COUNT(*)
FROM ft_attr
GROUP BY 1,2
ORDER BY 3 DESC;
```

Source	Campaign	Count
medium	interview-with-cool-tshirts-founder	622
nytimes	getting-to-know-cool-tshirts	612
buzzfeed	ten-crazy-cool-tshirts-facts	576
google	cool-tshirts-search	169

•2. What is the user journey? - continued

2.2 How many last touches is each campaign responsible for?

By modifying the previous query we can also obtain a last touch count. Here we use the **MAX timestamp** (rather than **MIN**) to ascertain the last user visit and which campaign and source motivated the last visit to the website.

Here we discover a larger variety of sources and campaigns than accounted for by first touches. This shows the leading source and campaign was the **email** of the **weekly-newsletter**, accounting for **447** last touches. **Medium**, with the '**interview with the CoolTShirts founder**' ,drops from first place in first touch count to sixth in the last touch count.

```
--Query to find last_touch count by campaign
using temporary tables

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

Source	Campaign	Count
email	weekly-newsletter	447
facebook	retargetting-ad	443
email	retargetting-campaign	245
nytimes	getting-to-know-cool-tshirts	232
buzzfeed	ten-crazy-cool-tshirts-facts	190
medium	interview-with-cool-tshirts-founder	184
google	paid-search	178
google	cool-tshirts-search	60

•2. What is the user journey? - continued

2.3 How many visitors make a purchase?

To find how many visitors make a purchase we can use a query to **COUNT** the number of **DISTINCT user_id** . By including a **WHERE** clause it is possible to restrict the results to only **COUNT** users who visited the **purchase page**. Here the data shows that **361** visitors to the website actually made a purchase, where first touch and last touch user_id counts are identical at **1,979**. Using this data we can calculate that the conversion rate of visitors into sales is **18.4 %** .

```
--Query to find distinct user_id that visited purchase page.

SELECT COUNT (distinct user_id)
FROM page_visits
WHERE page_name = '4 - purchase';
```

Count
361

•2. What is the user journey? - continued

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

By creating a temporary table, we can identify **user_id** and **last touches**, filtering the results using a **WHERE** clause to return only rows that had a last_touch on the '**4 – purchase**' page. A second temporary table adds source and campaign to this first set of information by **JOINing** the **page_visists** and **last_touch** tables on **user_id** and **timestamp**. Similar to the prior last touch SQL query the final **SELECT** statement **COUNTs** user id's with last touches on the purchase page where a source and campaign is associated with them..

Here we find that the **weekly-newsletter** campaign has the highest touch count on the purchase page, at **115**, while the **cool-tshirts-search** has the lowest, with a count of **2.**

```
--Query to find last_touch count by
campaign on the purchase page using
temporary tables.

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

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

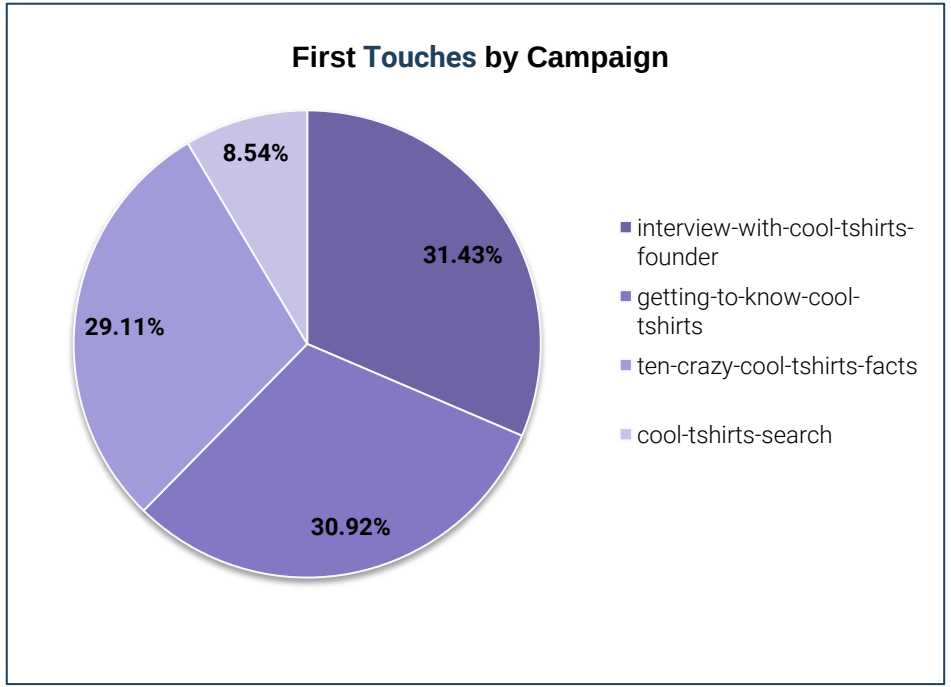
•2. What is the user journey? - continued

2.5a First visits to CoolTShirts website.

By looking at first touch, last touch and the last touch on the purchase page we can see that a typical user first clicks on the CoolTShirts site through one of three campaigns:

•interview-with-cool-tshirts-founder
•getting-to-know-cool-tshirts
•ten-crazy-cool-tshirts-facts : :

These three (out of a total of four) campaigns account for **91.46%** of first touches. However, as can be seen in further results obtained from last touch data we can see that these campaigns only account for **7%** of sales.



•2. What is the user journey? - continued

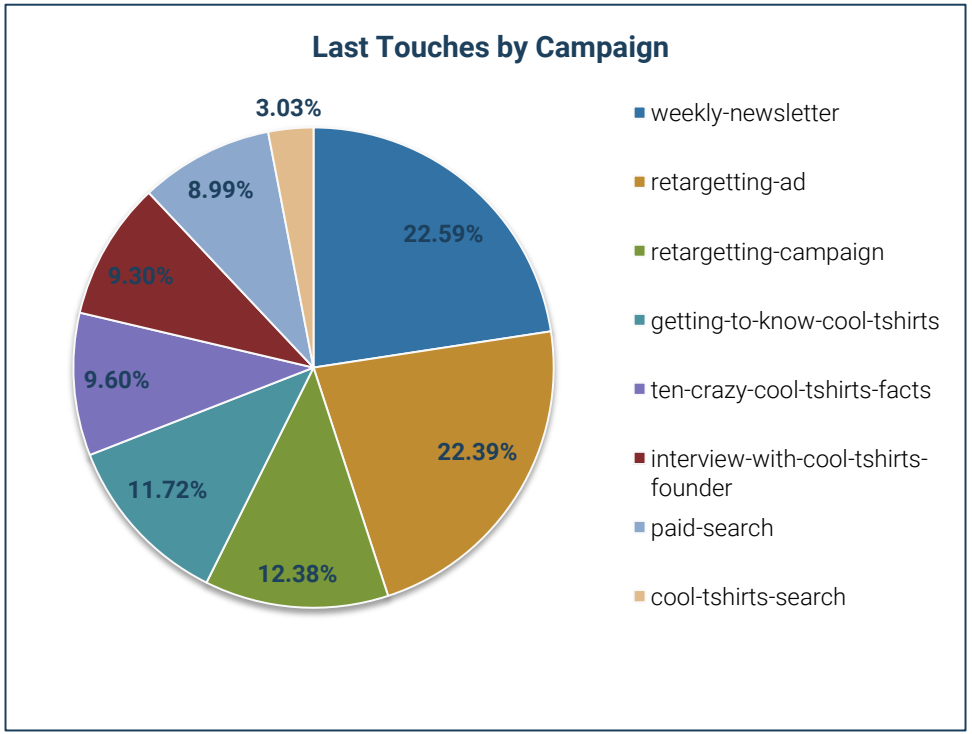
2.5b What the last touch information tell us.

Having analysed the first touch information we can now look at what brought prospective customers back to the website. Using this information along with the data obtained by querying the **COUNT** of last touches on the purchase page allows us to analyse which campaigns were most successful in prompting users to re-visit the site, and also which led to the most sales.

Having initially visited the site though one of the campaigns mentioned previously, the majority of returning users do so via a reminder or re-targeting ad. Here the data shows that the three most success campaigns leading to last touches are:

•weekly-newsletter
•retargetting-ad
•retargetting-campaign

These account for **57%** of returning users. We can see from the data that the three most successful first touch campaigns drop to **30.62%** of the total.



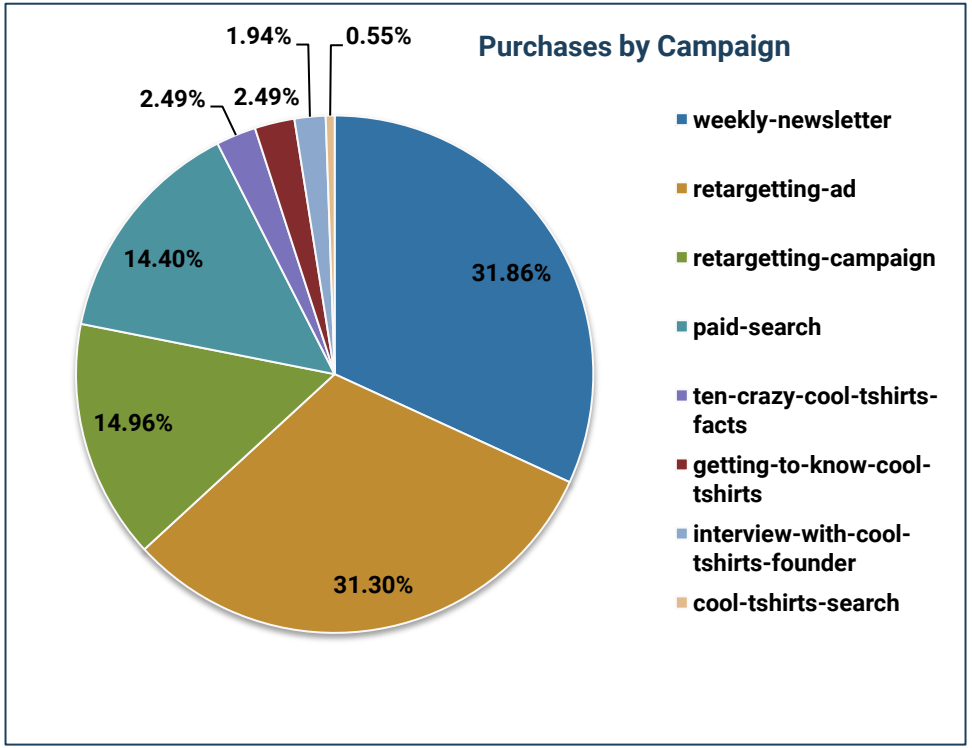
•2. What is the user journey? - continued

2.C Purchases

Looking at the purchasing data we can see that the campaigns responsible for the most first touches are the least effective at generating sales, with all four first touch campaigns accounting for only **7%** of sales. However, the three most effective last touch campaigns, which were all reminder campaigns - **weekly newsletter, retargeting ads, retargeting campaign -** account for **78% of sales.**

2.D Conclusion

From the data shown we can conclude that the typical user first clicks on the CoolTShirts website from one of three main campaigns, but that the 'reminder campaigns' are responsible for the majority of last touches and purchases. This indicates that most uses require a reminder to re-visit the CoolTShirts website. **18.4%** of these users then go on to make a purchase.



•3. Optimize the campaign budget

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

From the data we have obtained we can ascertain that three campaigns are responsible for the majority of first touches, these being the time a user first visits the site. Likewise there are three campaigns, all reminder or retargeting campaigns, that account for over 57.35% of last touch visits, and also for 78% of sales. This gives us six campaigns as the main focus of attention from which to select five for re-investment.

While the retargeting campaigns account for the majority of sales, and thus demonstrate their effectiveness in driving users to re-visit the site and to purchase items, they rely on an initial first touch from other campaigns to follow up on. Knowing this, my suggestion would be to re-invest in the three campaigns that generate the highest number of first touches as these provide the initial drive for most users to visit the site:

•interview-with-cool-tshirts-founder
•getting-to-know-cool-tshirts
•ten-crazy-cool-tshirts-facts :

These campaigns account for **91.46%** of first user visits to the CoolTShirts website. Re-investment should also focus on the two campaigns that perform highest with last touch counts and last touches on the purchase page:

•weekly-newsletter
•retargetting-ad

These account for **44.98%** of last touches, and **63.16%** of last touches on the purchase page. If funding was possible it would be prudent to re-invest in the **retargeting campaign** as well, something that might be achieved by cutting the poorest performing campaigns - the **cool-tshirts-search** and the **paid-search**. As can be seen in the table below both of these campaigns show a marked drop off in effectiveness when compared with the other campaigns promoted by CoolTShirts.

.

•3. Optimize the campaign budget – continued.

Campaign	First Touches	% First Touches	Last Touches	% Last Touches	Purchases	% Purchases
interview-with-cool-tshirts-founder	622	31.43%	184	9.30%	7	1.94%
getting-to-know-cool-tshirts	612	30.92%	232	11.72%	9	2.49%
ten-crazy-cool-tshirts-facts	576	29.11%	190	9.60%	9	2.49%
cool-tshirts-search	169	8.54%	60	3.03%	2	0.55%
weekly-newsletter	0	0.00%	447	22.59%	115	31.86%
retargetting-ad	0	0.00%	443	22.39%	113	31.30%
retargetting-campaign	0	0.00%	245	12.38%	54	14.96%
paid-search	0	0.00%	178	8.99%	52	14.40%