codecademy

Codeacademy Capstone

Learn SQL from Scratch

Jon Kalasz

2018-07-31

CoolTShirts: First- and Last-Touch Attribution

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 the Company

1. Get Familiar with the Company

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

- The source is the *medium* from which a campaign is ran. For example, the *email* source is used for the *newsletter* marketing campaign.

- A campaign is the marketing operation used to persuade potential buyers to visit CoolTShirts resulting in a purchase.

- Sources can be used for more than one campaign. In this table, for example, Google is the source for more than one campaign.

The code on the right will show us the number of sources, the number of campaigns, and which campaign is associated with its particular source. We can see the results for each query displayed on the next slide.

```
SELECT COUNT(DISTINCT utm_campaign) AS
'No_Campaigns'
FROM page_visits;



SELECT COUNT(DISTINCT utm_source) AS
'No_Sources'
FROM page_visits;



SELECT DISTINCT
    utm_campaign AS 'Campaign',
    utm_source AS 'Source'
FROM page_visits;
```

1. Get Familiar with the Company

When we execute the code from the previous slide, our results are shown as such:

```
SELECT COUNT(DISTINCT utm_campaign) AS
'No_Campaigns'
FROM page_visits;
```

No_Campaigns
8

```
SELECT COUNT(DISTINCT utm_source) AS
'No_Sources'
FROM page_visits;
```

No_Sources
6

```
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. Get Familiar with the Company

What pages are on their website?

We can run another query against the page_visits table to get the unique values for the page_name field:

```
SELECT DISTINCT
    page_name AS 'CTS Pages'
FROM page_visits;
```

CTS Pages
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

SELECT DISTINCT returns only unique, or different, values in the output. This eliminates any duplicates from the dataset when you are looking to see distinct records.

With this, we can see the unique pages in the CoolTShirts page_visits database.

What is the User Journey?

2. What is the User Journey?

How many **first-touches** is each campaign responsible for?

We can answer this question by first writing a query to define what a first-touch is based on the user ID and the minimum timestamp value associated with it.

From there, we can select the source and campaign while joining that dataset with the page_visits table and where the timestamps are equal. Once we've joined this data, we can select it along with the overall count of first-touches.

Thus, we can see the source, campaign, and number of first-touches it is responsible for:

Source	Campaign	First Touches
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

```
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 AS Source,
    ft_attr.utm_campaign AS Campaign,
    COUNT(*) AS 'First Touches'
FROM ft_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

2. What is the User Journey?

How many **last-touches** is each campaign responsible for?

We can answer this question by simply modifying our query for first-touches. Here, we want to select the largest (max) value for the timestamp as opposed to the smallest (min). Our *Common Table Expression* will represent last_touch rather than first_touch, and we will adjust our tables prefixes and count aliases to represent the new data output.

With these changes, we can observe the source, campaign, and number of last-touches it is responsible for:

Source	Campaign	Last Touches
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

```
WITH last_touch AS (
    SELECT
        user_id,
        MAX(timestamp) AS
last_touch_at
    FROM page_visits
    GROUP BY user_id
    ),
ft_attr 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
    )
SELECT
    ft_attr.utm_source AS Source,
    ft_attr.utm_campaign AS Campaign,
    COUNT(*) AS 'Last Touches'
FROM ft_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

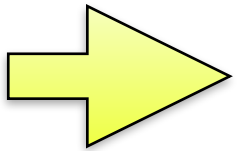
2. What is the User Journey?

How many customers make a purchase?

We can determine this number by selecting the unique, or **distinct**, user ID associated with the page_name of **4 - purchase**.

We add the **COUNT** function to our select statement to give us the amount of users who have purchased from CoolTShirts.

```
SELECT COUNT(DISTINCT user_id) AS
'Number Visitors that Purchase'
FROM page_visits
WHERE page_name = '4 - purchase';
```



Number Visitors that Purchase
361

2. What is the User Journey?

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

To determine this, we can go back to our last-touch query and filter our results based on the new criteria: last-touches only on the **purchase page**.

By adding this new criteria to our last_touch CTE and updating our COUNT alias, we will get the results we need:

Source	Campaign	Last Touches to Purchase
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

```
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
    ),
ft_attr 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
    )
SELECT
    ft_attr.utm_source AS Source,
    ft_attr.utm_campaign AS Campaign,
    COUNT(*) AS 'Last Touches to Purchase'
FROM ft_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

2. What is the User Journey?

What is the typical user journey?

We can observe that unique campaigns, eg. ‘interview with the founder’ or ’ten crazy facts’, draws the most interest and drives a majority of first-touches. This is good for introducing CoolTShirts to consumers, but is not responsible for very many purchases.

With the last query we ran against the purchase page, we can see that retargeting ads and newsletter campaigns were most effective at driving sales. While these retargeting campaigns drove the most purchases, the top first-touch and top-last touch campaigns probably compliment each other in order to finalize purchases.

Source	Campaign	First Touches
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

Source	Campaign	Last Touches to Purchase
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

Which Campaigns to Reinvest In?

3. Which Campaigns to Reinvest In?

With only five campaigns available to reinvest in, we should carefully review the data we've gathered. The fun and unique campaigns (shown below with first touches) may not lead immediately to a purchase, but do drive a significant amount of traffic and generates a ton of exposure for CoolTShirts.

With this, we can then focus on the weekly newsletter and the retargeting campaign via Facebook. If a user is signed up for our newsletter, that's guaranteed exposure the tends to lead to purchases, according to our data. And with over 2 billion monthly active users on Facebook, a retargeting campaign here tends to saturate enough users to keep purchases from this source high. The combination of these five campaigns will continue to contribute the most to purchases.

Source	Campaign	First Touches
medium	interview-with-cool-tshirts-founder	622
nytimes	getting-to-know-cool-tshirts	612
buzzfeed	ten-crazy-cool-tshirts-facts	576

+

Source	Campaign	Last Touches to Purchase
email	weekly-newsletter	115
facebook	retargetting-ad	113

* https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/