codecademy

Attribution

Learn SQL from Scratch

Lauren McKee

June 28th, 2018

Introduction to Attribution

- CoolTShirts is a T-shirt Company that markets their products via online advertisements and email ad campaigns.

- CoolTShirts uses *UTM parameters* to track their website visitors, and determine when visitors come to their website from an advertisement or remarketing email.

- This visitor data can then be aggregated and analyzed to evaluate the effectiveness of CoolTShirts' advertising campaigns, and allow them to shift & refocus their marketing investments where they will be the most successful.

Table of Contents

1. Familiarize Yourself With CoolTShirts

1.1 How Many Campaigns Does CoolTShirts Use?

The utm_campaigns column contains the names of the marketing campaigns that brought users to the CoolTShirts website. Marketing Campaigns are creative content (like articles, visual, or written advertisements) intended to inspire purchases or other actions from consumers. We can use the data in the utm_campaigns column to count the number of marketing campaigns in use by finding the number of unique values in the column.

To determine the number of campaigns CoolTShirts uses, count the number of distinct (unique) campaigns included in the utm_campaign column of the page_visits table, as shown in the query directly to the right.

The query at the right reveals that CoolTShirts has 8 distinct campaigns, as shown in the table, right.

```
SELECT COUNT(DISTINCT utm_campaign) AS number_of_distinct_campaigns
FROM page_visits;
```

number_of_distinct_campaigns
8

1.2 How Many Sources Does CoolTShirts Use?

UTM parameters allow CoolTShirts to collect data on where users navigated to their website from (the "source" of the web traffic); sources allow us to analyze ***where*** marketing campaigns are most successful. For example, the demographic composition of the audiences at sources like *The New York Times* & *Buzzfeed* is likely different, and we may want to consider this when deciding where to place ads & which ads to run on each source. Sources also include email campaigns (remarketing). We can use the data in the utm_source column to determine the number of sources by counting the number of unique values in this column.

To determine the number of sources CoolTShirts uses , count the number of distinct (unique) sources included in the utm_sources column of the page_visits table, as shown in the query directly to the right.

The query at the right reveals that CoolTShirts has 6 sources, as shown in the table, right.

```
SELECT COUNT(DISTINCT utm_source) AS number_of_distinct_sources
FROM page_visits;
```

number_of_distinct_sources
6

1.3 Which Source is Used for Each Campaign?

When evaluating the effectiveness of marketing campaigns, it will likely be helpful to keep in mind where (on what source) the campaign is being run.

To find the sources used for each campaign, we create a table that displays each unique campaign, and the source (location) where traffic associated with that campaign is coming from. This can be accomplished with the query displayed to the right. The columns were renamed for clarity.

The resulting table, right, lists each unique campaign in the “campaign” column. It lists the source where the campaign is being utilized in the “source” column.

```
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

1.4 What Pages Are On the CoolTShirts Website?

In order to assess the value of traffic generated from each campaign & source, it will be helpful to familiarize ourselves with the pages on the CoolTShirts website. We can use this information, for example, to analyze the percentage of traffic from each source resulting in a lead or a conversion (sale), among other useful analyses.

In order to familiarize ourselves with the pages on the CoolTShirts website, we can run the query (right) to create a list of the pages on the website.

The resulting table shows us there are 4 basic pages on the CoolTShirts website.

```
SELECT DISTINCT page_name AS pages
FROM page_visits;
```

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

2. The CoolTShirts User Journey

2.1 How Many First Touches is Each Campaign Responsible For?

We can determine the number of first touches (first website visits for a unique user) each campaign is responsible for using the query at the right, which:

1. Finds the first time each user visited the website (first_touch).
2. Joins this data with with a data set containing attribution information (ft_attribution).
3. Counts the number of first touches attributed to each campaign, and creates the table below.

In the table below, we learn that the interview-with-cool-tshirts-founder campaign is responsible for the most first touches (622), and that this campaign is run on the website ***medium***. This is followed by the getting-to-know-cool-tshirts campaign on the ***nytimes*** website (612 first touches), ten-crazy-cool-tshirts-facts on ***buzzfeed*** (576 first touches), and cool-tshirts-search on ***google*** (169 first touches).

source	campaign	number_of_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 1),
ft_attribution 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_attribution.utm_source AS source,
  ft_attribution.utm_campaign AS campaign,
  COUNT(*) as number_of_first_touches
FROM ft_attribution
GROUP BY 2
ORDER BY 3 DESC;
```

2.2 How Many Last Touches is Each Campaign Responsible For?

We can determine the number of last touches (most recent visit for a unique user) each campaign is responsible for using the query at the right, which:

1. Finds the last time each user visited the website (last_touch).
2. Joins this data with with a data set containing attribution information (lt_attribution).
3. Counts the number of last touches attributed to each campaign, and creates the table below.

In the table below, we learn that the weekly-newsletter campaign sent via email is responsible for the most last touches (447), followed by the retargetting-ad on ***facebook*** (443 last touches), retargetting-campaign sent via email (245), getting-to-know-cool-tshirts campaign on the ***nytimes*** website (232), ten-crazy-cool-tshirts-facts on ***buzzfeed*** (190), interview-with-cool-tshirts-founder on ***medium*** (184), ***google***'s paid-search (178), and cool-tshirts-search campaign on ***google*** (60).

source	campaign	number_of_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

```sql
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) AS last_touch_at
    FROM page_visits
    GROUP BY 1),
lt_attribution 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 lt_attribution.utm_source AS source,
  lt_attribution.utm_campaign AS campaign,
  COUNT(*) as number_of_last_touches
FROM lt_attribution
GROUP BY 2
ORDER BY 3 DESC;
```

2.3 How Many Visitors Make a Purchase?

We can determine how many unique visitors make a purchase by counting the number of unique visitors that visit the purchase page, using the query at right.

The results show that 361 unique visitors have made a purchase.

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

number_of_purchasing_vistors
361

2.4 How Many Last Touches *on the purchase page* is each campaign responsible for?

In order to fully understand the value of each marketing campaign, we can isolate how many last touches on the purchase page each campaign was responsible for (effectively, this is the number of purchases each campaign generated). This is accomplished by slightly modifying our query for last touches, by adding a "WHERE" clause to only count last touches on the purchase page.

The resulting table (below) shows us that CoolTShirt's remarketing campaigns seem to generate the most purchases, as we might expect – I know I might need a bit of a nudge to splurge on that super cool T-Shirt I found online last week. ☺

source	campaign	number_of_purchases
email	weekly-newsletter	114
facebook	retargetting-ad	112
email	retargetting-campaign	53
google	paid-search	52
nytimes	getting-to-know-cool-tshirts	9
buzzfeed	ten-crazy-cool-tshirts-facts	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
    GROUP BY 1),
lt_attribution 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_attribution.utm_source AS source,
  lt_attribution.utm_campaign AS campaign,
  COUNT(*) as number_of_purchases
FROM lt_attribution
WHERE lt_attribution.page_name = '4 - purchase'
GROUP BY 2
ORDER BY 3 DESC;
```

2.5.1 What is the Typical User Journey?

- Most users' first visit to the CoolTShirts website is inspired by perusing an article about CoolTShirts on popular news websites ***Medium*** (622 first touches), ***The New York Times*** (612 first touches), and ***Buzzfeed*** (576).
- A significant number of these users return to the site at a later date. We can see this by comparing the number of first touches and last touches attributed to each source. For example, 622 users' first touch on the site came from the medium interview-with-cool-tshirts-founder article, however only 184 users had their last touch with the site associated with this source. We can infer, therefore, that a significant number of users who visited the site from the medium article came back at some point (their last touch is associated with a different source).
- Many of the last touches are from remarketing/retargeting campaigns. The weekly-newsletter email accounts for 447 last touches, the Facebook retargeting-ad accounts for 443 last touches, and the email retargeting-ad accounts for 245.

source	campaign	number_of_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	number_of_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

source	campaign	number_of_purchases
email	weekly-newsletter	114
facebook	retargetting-ad	112
email	retargetting-campaign	53
google	paid-search	52
nytimes	getting-to-know-cool-tshirts	9
buzzfeed	ten-crazy-cool-tshirts-facts	9
medium	interview-with-cool-tshirts-founder	7
google	cool-tshirts-search	2

2.5.2 What is the Typical User Journey?

- If we do some simple math, we can see that the percentage of users who respond to remarketing with a purchase is quite high. Of the 447 last touches from the weekly newsletter, 114 users made a purchase (roughly 25.5%).

- Further analyses would be needed to confirm this absolutely, but with only the data generated in the course of the capstone project, I would expect

  the typical user journey to be article > leave site > targeting for remarketing > return to site > possible purchase

source	campaign	number_of_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	number_of_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

source	campaign	number_of_purchases
email	weekly-newsletter	114
facebook	retargetting-ad	112
email	retargetting-campaign	53
google	paid-search	52
nytimes	getting-to-know-cool-tshirts	9
buzzfeed	ten-crazy-cool-tshirts-facts	9
medium	interview-with-cool-tshirts-founder	7
google	cool-tshirts-search	2

3. Optimize The Campaign Budget

3.1.1 Which 5 Campaigns Should CoolTShirts Reinvest In?

- Because the typical user journey involves finding the website from a source like *Medium*, *The New York Times*, or *Buzzfeed*, leaving the website, and then being prompted to return for a possible purchase by a remarketing campaign (as demonstrated on a previous slide), I recommend keeping the 3 campaigns that are generating the most first touch traffic to the CoolTShirts website: interview-with-cool-tshirts-found (*Medium*), getting-to-know-cool-tshirts (*NYTimes*), ten-crazy-cool-tshirts-facts (*Buzzfeed*). These three campaigns account for 91.5% of first touches.

- The three remarketing campaigns are generating the most sales (weekly-newsletter, retargeting-ad, retargeting-campaign) (77.93% of purchases). I would recommend keeping two of these remarketing campaigns active, we can eliminate one of the two email remarketing campaigns because the email campaigns are likely serving the same (or a similar) list of contacts.

source	campaign	number_of_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	number_of_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

source	campaign	number_of_purchases
email	weekly-newsletter	114
facebook	retargetting-ad	112
email	retargetting-campaign	53
google	paid-search	52
nytimes	getting-to-know-cool-tshirts	9
buzzfeed	ten-crazy-cool-tshirts-facts	9
medium	interview-with-cool-tshirts-founder	7
google	cool-tshirts-search	2

3.1.2 Which 5 Campaigns Should CoolTShirts Reinvest In?

I would recommend eliminating the email remarketing campaign that costs more to run ***per purchase***. It's difficult to determine which campaign is more expensive per purchase without additional information about the labor involved in each, but assuming:

- The retargeting campaign is a drip campaign (a series of pre-written emails sent at pre-determined intervals after a user opts-in to receiving emails) and the weekly newsletter is fresh, new content every week, it likely costs considerably more in labor to send the weekly newsletter.
- The cost of the email services used to send these campaigns is the same.

Based on this line of thought, I recommend keeping the retargeting-ad (facebook) and the retargeting-campaign (email) active.

3.1.3 Which 5 Campaigns Should CoolTShirts Reinvest In?

In Summary, I would recommend the following 5 campaigns remain active:

- Interview-with-cool-tshirts-founder (***Medium***)
- Getting-to-know-cool-tshirts (***NYTimes***)
- Ten-crazy-cool-tshirts-facts (***Buzzfeed***)
- Retargeting-ad (***Facebook***)
- Retargeting-campaign (email)