codecademy

Attribution Queries

Learn SQL from Scratch
Nolan Padgett Jr
6/25/19

Example Table of Contents

1. Get familiar with CoolTShirts
2. What is the user journey?
3. Optimize the campaign budget

Get Familiar with CoolTShirts

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

This question was broken down into three separate queries to provide an understanding of our data set:
- Our first query was a simple distinct count that was able to return the number of campaigns based upon the unique campaign names.
- The result of the query being 8 distinct campaigns.

```
SELECT COUNT(DISTINCT utm_campaign) AS 'Number of Campaigns'
FROM page_visits;


SELECT COUNT(DISTINCT utm_source) AS 'Number of Sources'
FROM page_visits;


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

Number of Campaigns
8

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

This question was broken down into three separate queries to provide an understanding of our data set:

- Our second query was a simple distinct count that was able to return the number of sources based upon the unique source names. Similar to the query returning the number of campaigns, with slightly different values.
- The result of the query being 6 distinct sources.

```
SELECT COUNT(DISTINCT utm_campaign) AS 'Number of
Campaigns'
FROM page_visits;


SELECT COUNT(DISTINCT utm_source) AS 'Number of
Sources'
FROM page_visits;


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

Number of Sources
6

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

This question was broken down into three separate queries to provide an understanding of our data set:
- Our third query was pulled each of our distinct campaigns in one column of our resulting table, and then a column that reported their corresponding source based on our dataset.
- The result of the query is shown in the table below.

```
SELECT COUNT(DISTINCT utm_campaign) AS 'Number of
Campaigns'
FROM page_visits;


SELECT COUNT(DISTINCT utm_source) AS 'Number of
Sources'
FROM page_visits;


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

What pages are on their website?

- The question regarding the pages on the site was answered by querying the dataset for each of the distinct page names in our data set.
- The result of the query is shown in the table below.

```
SELECT DISTINCT page_name AS 'Page Names'
FROM page_visits;
```

Page Names
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

What is the User Journey?

How many first touches is each campaign responsible for?

To establish the number of first touches, a temporary table called first_touch was created to find the earliest date that each individual user was on the site and then joined with our primary dataset.

The result of the query is shown in the table below.

```
WITH first_touch AS (
    SELECT user_id,
        MIN(timestamp) as first_touch_at
    FROM page_visits
    GROUP BY user_id)
SELECT pv.utm_campaign, pv.utm_source, COUNT(*)
FROM first_touch ft
JOIN page_visits pv
    ON ft.user_id = pv.user_id
    AND ft.first_touch_at = pv.timestamp
GROUP BY 1
ORDER BY 1;
```

utm_campaign	utm_source	COUNT(*)
cool-tshirts-search	google	169
getting-to-know-cool-tshirts	nytimes	612
interview-with-cool-tshirts-founder	medium	622
ten-crazy-cool-tshirts-facts	buzzfeed	576

How many last touches is each campaign responsible for?

To establish the number of first touches, a temporary table called last_touch was created to find the latest date that each individual user was on the site and then joined with our primary dataset.

The result of the query is shown in the table below.

```
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) as last_touch_at
    FROM page_visits
    GROUP BY user_id)
SELECT pv.utm_campaign, pv.utm_source, COUNT(*)
FROM last_touch AS lt
JOIN page_visits AS pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
GROUP BY 1
ORDER BY 1;
```

utm_campaign	utm_source	COUNT(*)
cool-tshirts-search	google	60
getting-to-know-cool-tshirts	nytimes	232
interview-with-cool-tshirts-founder	medium	184
paid-search	google	178
retargetting-ad	facebook	443
retargetting-campaign	email	245
ten-crazy-cool-tshirts-facts	buzzfeed	190
weekly-newsletter	email	447

How many visitors make a purchase?

The number of visitors who make a purchase is found by filtering our data set down to only those transactions that made it to the purchase stage with a WHERE clause, and then using a count of all the distinct user ids left in that result.

The result of the query being 361 customers.

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

COUNT(DISTINCT user_id)
361

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

By taking a combination of concepts in two previous queries we are able to draw the answer to the last touches on the purchase page. By taking the temporary last_touch query from before and using the same filtering concept we can determine the campaign responsible for the last touches on each purchase.

The result of the query is shown in the table below.

```
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) as last_touch_at
    FROM page_visits
            WHERE page_name IS '4 - purchase'
    GROUP BY user_id)
SELECT pv.utm_campaign, pv.utm_source, COUNT(*)
FROM last_touch AS lt
JOIN page_visits AS pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
GROUP BY 1
ORDER BY 1;
```

utm_campaign	utm_source	COUNT(*)
cool-tshirts-search	google	2
getting-to-know-cool-tshirts	nytimes	9
interview-with-cool-tshirts-founder	medium	7
paid-search	google	52
retargetting-ad	facebook	113
retargetting-campaign	email	54
ten-crazy-cool-tshirts-facts	buzzfeed	9
weekly-newsletter	email	115

What is the typical user journey?

1. The majority of users first find the site through the Getting to know CoolTShirts, Interview with CoolTShirts Founder, and Ten Crazy CoolTShirts campaigns, yet they do not tend to complete their purchases from those visits.
2. The typical user will come back to the site at a later time after their initial contact through the Weekly Newsletter and the Retargeting Campaign in which they will then make their decision on whether or not they will actually purchase, in which the more that 75% of the customers who make purchases actually come through these two avenues.

First Touches		Last Touches		Last Touches On Purchases	
Campaign	#	Campaign	#	utm_campaign	#
cool-tshirts-search	169	cool-tshirts-search	60	cool-tshirts-search	2
		getting-to-know-cool-tshirts	232	getting-to-know-cool-tshirts	9
getting-to-know-cool-tshirts	612	interview-with-cool-tshirts-founder	184	interview-with-cool-tshirts-founder	7
		paid-search	178	paid-search	52
interview-with-cool-tshirts-founder	622	retargetting-ad	443	retargetting-ad	113
		retargetting-campaign	245	retargetting-campaign	54
		ten-crazy-cool-tshirts-facts	190	ten-crazy-cool-tshirts-facts	9
ten-crazy-cool-tshirts-facts	576	weekly-newsletter	447	weekly-newsletter	115

Optimize the Campaign Budget

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

I would invest in the following 5 campaigns if I were CoolTShirts:
1. Getting to know CoolTShirts
2. Interview with CoolTShirts Founder
3. Ten Crazy CoolTShirts Facts
4. Weekly Newsletter
5. Retargeting Campaign

I would reinvest in these campaigns for two sets of reasons:
1. I would first re-invest in the Getting to know CoolTShirts, Interview with CoolTShirts Founder, and Ten Crazy CoolTShirts Facts campaigns, they each do the best job of bringing customers to your site. Peaking their interest in the company
2. I would then re-invest in the Weekly Newsletter and the Retargeting Campaign, because based on the queries I created during this project, they play the biggest part in helping to close the deal for CoolTShirts. They are not only responsible for the most last touches by far, but they are also responsible for the most last touches on the purchase page.