codecademy

Attribution Queries

Learn SQL from Scratch
Anil Mallya
1/2/19

Table of Contents

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

1. Getting Familiar with CoolTShirts

CoolTShirts Ad Campaigns

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

Q. **How many campaigns and sources does cooltshirts have?**
A.   The first two queries on the right show that cooltshirts has 8 campaigns from 6 different sources.

Q. **How are the campaigns related?**
A.   We can see from the result of the third query (on the left) that cooltshirts has two email campaigns and two google campaigns in addition to one campaign each on buzzfeed, facebook, medium and the new york times.

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

SELECT COUNT(DISTINCT(utm_source))
FROM page_visits;

SELECT DISTINCT utm_source, utm_campaign
FROM page_visits
     ORDER BY utm_source ASC;
```

What distinct pages can be found on CoolTShirts Website?

If we select the distinct page names from the page_visits table, we can see that cooltshirts has 4 pages. The results of this query are shown on the right.

```
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?

How many "first touches" is each campaign responsible for?

user_id	first_touch_at	utm_source	utm_campaign	count(utm_campaign)
99990	2018-01-13 23:30:09	medium	interview-with-cool-tshirts-founder	622
99933	2018-01-25 00:04:39	nytimes	getting-to-know-cool-tshirts	612
99765	2018-01-04 05:59:46	buzzfeed	ten-crazy-cool-tshirts-facts	576
99684	2018-01-13 13:20:49	google	cool-tshirts-search	169

Utilizing the first touch attribution query that we created previously, we can determine how many first touches exist. It appears that the campaign on Medium generated the most first touches.

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

How many "last touches" is each campaign responsible for?

user_id	last_touch_at	utm_source	utm_campaign	last_touches
99933	2018-01-26 06:18:39	email	weekly-newsletter	447
99928	2018-01-24 05:26:09	facebook	retargetting-ad	443
99990	2018-01-16 11:35:09	email	retargetting-campaign	245
99589	2018-01-15 04:55:43	nytimes	getting-to-know-cool-tshirts	232
99765	2018-01-04 05:59:47	buzzfeed	ten-crazy-cool-tshirts-facts	190
99838	2018-01-02 07:40:34	medium	interview-with-cool-tshirts-founder	184
98840	2018-01-10 04:58:48	google	paid-search	178
99344	2018-01-18 21:36:32	google	cool-tshirts-search	60

By altering the first touch attribution query that we created, we can determine the last touches. We look for the max of the timestamp from the page views table. See the code below for full details. The weekly newsletter generated the most last touches.

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

3. Optimize the Campaign Budget

How Many Visitors make a Purchase?

From the query to the right we see that 361 distinct visitors made purchases.

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

purchases
361

Which campaigns are responsible for last touches that lead to purchases?

user_id	last_touch_at	utm_source	utm_campaign	last_touches
99933	2018-01-26 06:18:39	email	weekly-newsletter	115
99897	2018-01-06 09:41:19	facebook	retargetting-ad	113
99285	2018-01-24 09:00:58	email	retargetting-campaign	54
94567	2018-01-19 16:37:58	google	paid-search	52
92172	2018-01-16 15:15:29	nytimes	getting-to-know-cool-tshirts	9
98651	2018-01-15 04:17:36	buzzfeed	ten-crazy-cool-tshirts-facts	9
83547	2018-01-10 18:20:21	medium	interview-with-cool-tshirts-founder	7
95650	2018-01-18 00:25:00	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
            )
SELECT lt.user_id,
    lt.last_touch_at,
    pv.utm_source,
    pv.utm_campaign,
    Count(utm_campaign) AS last_touches
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
GROUP BY utm_campaign
ORDER BY COUNT(utm_campaign) DESC;
```

By adding the where statement to our last touch query, we can determine which last touches were on the purchase page. The weekly email newsletter was responsible for the most purchases, followed by the facebook retargetting ads. This mirrors what we saw with the last touch query previously.

4. Conclusions and Summary

Final Thoughts

From the data collected, we can see that while the ny times and medium campaigns generated a lot of site traffic, they did not lead to a significant number of purchases. On the other hand, Facebook and Paid Google Search were able to achieve a much higher percentage purchase conversion, and should be a focus for our advertising efforts.