codecademy

Attribution Queries - Capstone

Learn SQL from Scratch

Erwin C. Palma

December 17, 2018

Table of Contents

1. Getting Familiar w/ The Company
2. Finding A User’s Journey
3. Optimizing The Campaign Budget

1. Getting Familiar w/ The Company

1.1 Getting To Know CoolTShirts

My first task involved understanding how many unique ad campaigns and sources CoolTShirts uses, as well as to find a connection between the campaigns and sources

- The campaigns, or **utm_campaigns**, are the types of ads run on several different websites. The sources, or **utm_source**, are the websites that customers clicked each corresponding
- Using the SQL commands shown on the right, I learned that there are **8** unique ad campaigns, and **6** unique sources

Campaign
8

Source
6

```
select count(DISTINCT utm_campaign) AS 'Campaign'
from page_visits;

select count(DISTINCT utm_source) AS 'Source'
from page_visits;
```

1.2 Getting To Know CoolTShirts

Next, I tried to find the connection between the campaigns and sources

- Using the SQL command below, I was able to find out which ad campaigns ran on which websites that customers clicked on to get to the CoolTShirts websites
- Except for the “Google” and “email” source, each source had one CoolTShirts ad campaign featured on their website. Those findings are shown to the right.

```
select DISTINCT utm_source, utm_campaign
from page_visits;
```

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

2. Finding A User’s Journey

2.1 Finding A User's Journey

To understand the path a typical user makes once clicking on a campaign, I first had to find out how many first touches each campaign was responsible for.

- In my first-touch query,, I created a table to count first touches. I created then joined a temporary table named **first_touch** together with the **page_visits** table at two different columns to find the data I was looking for. The table below shows my findings, with the **COUNT (utm_campaign)** column giving me the data on how many first touches each campaign brought in.

```
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 5 DESC;
```

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

2.2 Finding A User's Journey

Finding out how many last touches each campaign was responsible for was similar to the way I built the first touch table.

- The structure of the query this time around was almost the same as last slide's query. This time I created then joined a temporary table called **last_touch** together with the **page_visits** table at two different columns. The screenshot below shows my findings, with the **COUNT (utm_campaign)** column giving me the data on how many last touches each campaign brought in.

```
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)
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 5 DESC;
```

user_id	last_touch_at	utm_source	utm_campaign	COUNT (utm_campaign)
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

2.3 Finding A User's Journey

After figuring out how many first touches and last touches each campaign generated, I wanted to find out how many visitors to CoolTShirts actually make a purchase.

- I ran a query where I counted each distinct user_id from the page_visits table where the user reached the 'purchase' page. I was able to find out that in our data, **361 visitors** made a purchase.

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

2.4 Finding A User's Journey

Having found out that there were 361 visitors who made a purchase, I wanted to see how many last touches on the purchase page each campaign was responsible for.

- Using the previous query I made to find the last touch distribution for each campaign, I added a **WHERE** clause that filtered rows where each campaign led to a purchase from a last touch.
- The **Last_Touch_To_Purchase** column shows how many last touches on the purchase page each **Campaign** was responsible for.

```
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 AS 'Campaign',
    COUNT (utm_campaign) AS
'Last_Touch_To_Purchase'
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 5 DESC;
```

Campaign	Last_Touch_To_Purchase
weekly-newsletter	115
retargetting-ad	113
retargetting-campaign	54
paid-search	52
getting-to-know-cool-tshirts	9
ten-crazy-cool-tshirts-facts	9
interview-with-cool-tshirts-founder	7
cool-tshirts-search	2

2.5 Finding A User's Journey

To be even more complete, I decided to find out where the other users were last touching the website if it was not on the *purchase* page.

- I edited the query I used to find all the last touches on the *purchase* page to count all the last touches on both the *shopping_cart* page and the *checkout* page. The two queries to the right showcase the total that made it to the *shopping_cart* page and *checkout* page respectively.
- This yielded results that of all the users who first found the CoolTShirts, **1881** users last used the *shopping_cart* page and **1431** users last used the *checkout* page.

```
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) as last_touch_at
    FROM page_visits
      WHERE page_name = '2 - shopping_cart'
    GROUP BY user_id)
SELECT COUNT(DISTINCT pv.user_id)
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp;
```

```
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) as last_touch_at
    FROM page_visits
      WHERE page_name = '3 - checkout'
    GROUP BY user_id)
SELECT COUNT(DISTINCT pv.user_id)
FROM last_touch lt
JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp;
```

2.6 Finding A User's Journey

Using all the data I was able to organize with the *page_visits* table, I then asked "what is a typical user's journey through the website?" If the overall goal of each ad campaign is to lead a visitor to make a purchase on CoolTShirts, then breaking down the data and analyzing how these campaigns did would detail how it is able to reach consumers and what a user's journey looks like.

- Altogether, **1979** users first visited the CoolTShirts website via any of their ad campaigns.
- Out of these 1979 users, **361** users' last touches on the website were made to purchase merchandise from the store.
- Furthermore, **1881** users found their way to the *shopping_cart* page and **1431** users made their way to the *checkout* page.
- After dividing how many users first visited the website with how many users actually made purchases, you see that **18.2%** of users who found the website for the first time eventually ended up making a purchase from CoolTShirts.
- **72.3%** make it all the way to the third funnel, the *checkout* page. However, there is a drastic drop off from the *checkout* page to the *purchase* page, suggesting that for some reason most users don't seem to pull the trigger and make a purchase.

3. Optimizing The Campaign Budget

3.1 Optimizing The Campaign Budget

Whether or not CoolTShirts finds the ad campaigns effective based on the numbers I calculated, it is clear that some campaigns were more effective than others. If CoolTShirts choose to reinvest or rework the campaigns, these five would be my picks.

1. The first two campaigns I would choose to invest more in are the *weekly-newsletter* and *retargetting-ad* from the email and Facebook sources. These two are far and away the most effective ad campaigns that lead visitors to making a purchase accounting for **238** of the 361 purchases. Reinvesting more resources into these campaigns by making these campaigns more visible would prove wise to help increase purchases even more.
2. The last three are the *interview-with-cool-tshirts-founder*, *getting-to-know-cool-tshirts,* and *ten-crazy-cool-tshirt-facts* campaigns. They are the top three campaigns that users clicked on to first get to CoolTShirts, yet they are in the bottom four in the **Last_Touch_To_Purchase** category. Combined, they make up for **1810** of the 1979 first touches into the website, yet they make up for **25** of the 361 purchases. While clearly effective in bringing users to the site, the campaign needs to be overhauled to bridge the huge gap in purchases that these three campaigns have with respect to the top campaigns.

utm_source	Campaign	Last_Touch_To_Purchase
email	weekly-newsletter	115
facebook	retargetting-ad	113

utm_campaign	First_touch_numbers
interview-with-cool-tshirts-founder	622
getting-to-know-cool-tshirts	612
ten-crazy-cool-tshirts-facts	576