codecademy

# Capstone Project - Attribution Queries

Learn SQL from Scratch

Kane Wallsmith

to_date('23–Apr–2019' , 'dd-Mon-yyyy')

# Table of Contents

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

# Contents of Table

(Database Schema)

| page_visits<br>5692 rows | |
|---|---|
| **page_name** | TEXT |
| **timestamp** | TEXT |
| **user_id** | INTEGER |
| **utm_campaign** | TEXT |
| **utm_source** | TEXT |

| page_name | timestamp | user_id | utm_campaign | utm_source |
|---|---|---|---|---|
| 1 - landing_page | 2018-01-24 03:12:16 | 10006 | getting-to-know-cool-tshirts | nytimes |
| 2 - shopping_cart | 2018-01-24 04:04:16 | 10006 | getting-to-know-cool-tshirts | nytimes |
| 3 - checkout | 2018-01-25 23:10:16 | 10006 | weekly-newsletter | email |
| 1 - landing_page | 2018-01-25 20:32:02 | 10030 | ten-crazy-cool-tshirts-facts | buzzfeed |
| 2 - shopping_cart | 2018-01-25 23:05:02 | 10030 | ten-crazy-cool-tshirts-facts | buzzfeed |
| 3 - checkout | 2018-01-28 13:26:02 | 10030 | retargetting-campaign | email |
| 4 - purchase | 2018-01-28 13:38:02 | 10030 | retargetting-campaign | email |
| 1 - landing_page | 2018-01-05 18:31:17 | 10045 | getting-to-know-cool-tshirts | nytimes |
| 2 - shopping_cart | 2018-01-05 21:16:17 | 10045 | getting-to-know-cool-tshirts | nytimes |
| 3 - checkout | 2018-01-09 03:05:17 | 10045 | retargetting-ad | facebook |

# 1. Get Familiar With the Company

# 1.1 Campaigns and Sources and Pages, oh my!

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

- There are **8** distinct Campaigns and **6** distinct Sources
- Campaigns and Sources are separate but related. A campaign is a specific ad or email targeted with a specific audience or product in mind. Each Campaign will have a Source, like a website or communication method, which identifies how the traffic was sent to the product page

What pages are on their Website?

- There are **4** distinct pages within the CoolTshirts database
- Each page corresponds with a specific 'step' in the customer journey to making a purchase
- Each campaign leads users to one of the 4 pages on the website (Typically landing page or checkout)

| Campaign_Name |
| --- |
| getting-to-know-cool-tshirts |
| weekly-newsletter |
| ten-crazy-cool-tshirts-facts |
| retargetting-campaign |
| retargetting-ad |
| interview-with-cool-tshirts-founder |
| paid-search |
| cool-tshirts-search |

| Source_Name |
| --- |
| nytimes |
| email |
| buzzfeed |
| facebook |
| medium |
| google |

| Page_Name |
| --- |
| 1 - landing_page |
| 2 - shopping_cart |
| 3 - checkout |
| 4 - purchase |

```sql
SELECT
  COUNT(DISTINCT utm_campaign) AS 'campaign_count',
  COUNT(DISTINCT utm_source) AS 'source_count'
FROM page_visits;


SELECT
  DISTINCT utm_campaign AS 'Campaign_Name',
           utm_source AS 'Source_Name',
           page_name AS 'Page_Name'
FROM page_visits;
```

# 2. What is the User Journey?

# 1.1 No touchy!

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

- First touch attribution is essentially just the first touchpoint/interaction a user has with the website. It can be analyzed by applying a MIN function to the timestamp of each user_id, effectively showing the first date and time a particular user visited the page. So if a customer discovers a website via a certain campaign, that channel would get the credit of first_touch. It lets businesses know what is bringing customers in the door.

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

- Last touch attribution is the opposite of first touch. Applying the MAX aggregate function on the user_id timestamp will yield the last touchpoint. This analysis provides data insight into which channel is responsible for bringing customers to the final purchasing stages. Businesses use this metric to understand which campaigns are more effective at producing purchases.

| Campaign_Name | Source | First_Touches | Last_Touches |
|---|---|---|---|
| getting-to-know-cool-tshirts | nytimes | 612 | 232 |
| weekly-newsletter | email | NULL | 447 |
| ten-crazy-cool-tshirts-facts | buzzfeed | 576 | 190 |
| retargetting-campaign | email | NULL | 245 |
| retargetting-ad | facebook | NULL | 443 |
| interview-with-cool-tshirts-founder | medium | 622 | 184 |
| paid-search | google | NULL | 178 |
| cool-tshirts-search | google | 169 | 60 |
| TOTAL | | 1979 | 1979 |

```sql
WITH first_touch AS (
--WITH last_touch AS
    SELECT user_id,
        MIN(timestamp) AS first_touch_at
        --MAX(timestamp) AS last_touch_at
    FROM page_visits
    GROUP BY user_id),
ft_attr AS (
  SELECT ft.user_id,
         ft.first_touch_at,
         pv.utm_source 'Source',
         pv.utm_campaign 'Campaign_Name'
  FROM first_touch ft
  JOIN page_visits pv
    ON ft.user_id = pv.user_id
    AND ft.first_touch_at = pv.timestamp)
--lt_attr AS (
--SELECT lt.user_id,
          --lt.first_touch_at,
          --pv.utm_source 'Source',
          --pv.utm_campaign 'Campaign_Name'
  --FROM last_touch lt
  --JOIN page_visits pv
    --ON lt.user_id = pv.user_id
    --AND lt.last_touch_at = pv.timestamp)
SELECT Source,
       Campaign_Name,
       COUNT(*) AS 'First_Touches'
       --COUNT(*) AS 'Last_Touches'
FROM ft_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

# 1.2 Its all about the Ba-bling

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

- Using the same construct from the last_touch query, adding in a WHERE clause that filters the results of the last_touch temporary table, will allow us to focus on ONLY the last touches associated with the 'purchase' page.

How many visitors make a purchase?

- By adding up the PP_Last_Touches column ( or simply querying the number of user_id associated with the page_name '4 – purchase' ) the total visitor purchases comes to **361**

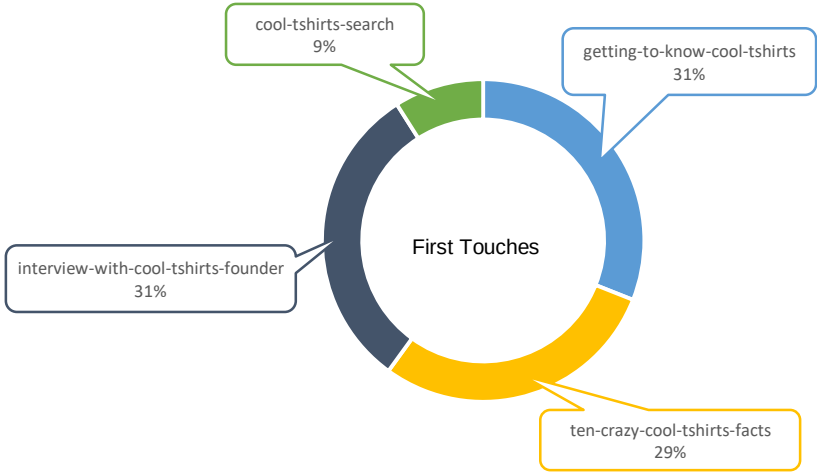
| Campaign_Name | Source | Last_Touches | PP_Last_Touches |
|---|---|---|---|
| getting-to-know-cool-tshirts | nytimes | 232 | 9 |
| weekly-newsletter | email | 447 | 115 |
| ten-crazy-cool-tshirts-facts | buzzfeed | 190 | 9 |
| retargetting-campaign | email | 245 | 54 |
| retargetting-ad | facebook | 443 | 113 |
| interview-with-cool-tshirts-founder | medium | 184 | 7 |
| paid-search | google | 178 | 52 |
| cool-tshirts-search | google | 60 | 2 |
| TOTAL | | 1979 | 361 |

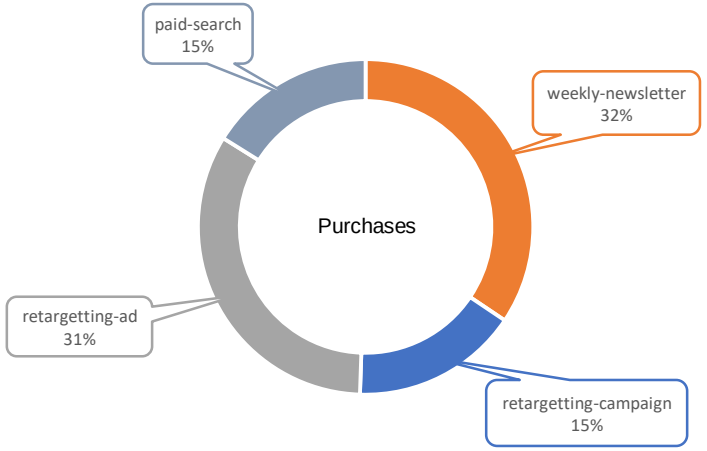
```sql
WITH last_touch AS (
    SELECT user_id,
        MAX(timestamp) AS last_touch_at
    FROM page_visits
    WHERE page_name LIKE '4%'
    GROUP BY user_id),
lt_attr AS (
SELECT lt.user_id,
       lt.last_touch_at,
       pv.utm_source 'Source',
       pv.utm_campaign 'Campaign_Name'
  FROM last_touch lt
  JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp)
SELECT Source,
       Campaign_Name,
       COUNT(*) AS 'Last_Touches'
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3 DESC;

```

# 1.3 But really, its about the journey

What is the typical user journey?

- The typical user journey would involve stumbling upon CoolTShirts through one of the following campaigns: cool-tshirts-search, getting-to-know-cool-tshirts, interview-with-cool-tshirts-founder, or ten-crazy-cool-tshirts-facts.
- The next user step seems to be completely forgetting about that awesome new t-shirt you just found until.....
- The user is reminded about that week-old shopping cart by one of the following campaigns: retargetting-ad, retargetting-campaign, weekly-newsletter or paid-search



# 3. Optimize the Campaign Budget

# 1.3 Its not about the result, its about the journey

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

Based on the data table below, there are a few key take-aways for consideration:

**1 –** Some campaigns are more effective at getting customers 'in the door' while others provide a higher Last_touch and Purchase percentage. Using a combination of the two would allow for the most encompassing campaign re-investment strategy.
**2 –** Rows highlighted below in green have been chosen based on their higher first_touch percentage or their purchases percentage.
**3 –** The selected combination of campaigns would allow CoolTShirts to maintain 91% of its current first_touch rates and 69% of the purchase rate

- One caveat would be the ROI of each campaign. Because this data was not included in the data set, it is impossible to tell how efficient individual campaigns are based on their price. For example if the 'retargeting-campaign' total cost was only $50, but the 'ten-crazy-cool-tshirts-facts' had a sticker price of $10,000, it may be more cost effective to run with the retargeting-campaign. (plus buzzfeed is total garbage, but that's just my personal opinion)

| Campaign_Name | Source | First_Touches | % of Total | Last_Touches | % of Total | Purchases | % of Total |
|---|---|---|---|---|---|---|---|
| getting-to-know-cool-tshirts | nytimes | 612 | 31% | 232 | 12% | 9 | 2% |
| weekly-newsletter | email | NULL | 0% | 447 | 23% | 115 | 32% |
| ten-crazy-cool-tshirts-facts | buzzfeed | 576 | 29% | 190 | 10% | 9 | 2% |
| retargetting-campaign | email | NULL | 0% | 245 | 12% | 54 | 15% |
| retargetting-ad | facebook | NULL | 0% | 443 | 22% | 113 | 31% |
| interview-with-cool-tshirts-founder | medium | 622 | 31% | 184 | 9% | 7 | 2% |
| paid-search | google | NULL | 0% | 178 | 9% | 52 | 15% |
| cool-tshirts-search | google | 169 | 9% | 60 | 3% | 2 | 1% |
| TOTAL | | 1979 | 100% | 1979 | 100% | 361 | 100% |