



Marketing Attributions

Analyze Data with SQL

Sergio Ventura Fiorentino Recabarren

16 september 2022

Introduction.

UTM parameters are a way to track visits to a website. Developers, marketers, and analysts use them to capture information such as time, attribution source, and attribution medium for each user visit.

page_visits	
A table describing each time a user visits the CoolTShirts website	
Column	Description
user_id	A unique identifier for each visitor to a page
timestamp	The time at which the visitor came to the page
page_name	The title of the section of the page that was visited
utm_source	Identifies which site sent the traffic (i.e., google, newsletter, or facebook_ad)
utm_campaign	Identifies the specific ad or email blast (i.e., june-21-newsletter or memorial-day-sale)

Introduction

- First touch attribution only considers the first source for each customer. This is a good way to find out how visitors initially discover a website.
- Last touch attribution only considers the last source for each customer. This is a good way to find out how visitors return to a website, especially to make a final purchase.

Introduction.

CoolTShirts sells t-shirts of all kinds, as long as they're T-shaped and cool. Recently, CTS started some marketing campaigns to increase website visits and purchases. Using touch attribution, they would like to map their customer's journey: from initial visit to purchase. They can use that information to optimize their marketing campaigns. CoolTShirts can reinvest in 5 campaigns.

With the results obtained in this work, it will be decided which are the 5 campaigns in which it will be reinvested.

1. Get familiar with CoolTShirt

1.1 Take a look at the data, types, number of rows and columns.

Query Results				
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
Database Schema				
page_visits				
name			type	
page_name			TEXT	
timestamp			TEXT	
user_id			INTEGER	
utm_campaign			TEXT	
utm_source			TEXT	
Rows: 5692				

```
-- You can put your query here
SELECT *
FROM page_visits
LIMIT 10;
```

1.2 How many campaigns and sources does CoolTShirts use? Which source is used for each campaign?

Use three queries:

one for the number of distinct campaigns.

one for the number of distinct sources.

one to find how they are related.

Query Results	
COUNT(DISTINCT utm_campaign)	
8	
COUNT(DISTINCT utm_source)	
6	
utm_campaign	utm_source
getting-to-know-cool-tshirts	nytimes
weekly-newsletter	email
ten-crazy-cool-tshirts-facts	buzzfeed
retargeting-campaign	email
retargeting-ad	facebook
interview-with-cool-tshirts-founder	medium
paid-search	google
cool-tshirts-search	google

```
-- You can put your query here
SELECT COUNT(DISTINCT utm_campaign)
FROM page_visits;
```

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

```
SELECT DISTINCT utm_campaign, utm_source
FROM page_visits;
```

1.3 What pages are on the CoolTShirts website?

Find the distinct values of the page_name column.

```
-- You can put your query here
SELECT DISTINCT page_name
FROM page_visits;

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

Query Results

page_name
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase
COUNT(DISTINCT page_name)
4

2. What is the user journey

2.1 How many first touches is each campaign responsible for?

```
-- You can put your query here
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;
```

Query Results				
user_id	first_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
10048	2018-01-16 04:17:46	medium	interview-with-cool-tshirts-founder	622
10006	2018-01-24 03:12:16	nytimes	getting-to-know-cool-tshirts	612
10030	2018-01-25 20:32:02	buzzfeed	ten-crazy-cool-tshirts-facts	576
10925	2018-01-20 10:21:30	google	cool-tshirts-search	169

2.2 How many **last touches** is each campaign responsible for?

```
-- You can put your query here
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;
```

Query Results

user_id	last_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
10006	2018-01-25 23:10:16	email	weekly-newsletter	447
10045	2018-01-09 03:05:17	facebook	retargeting-ad	443
10030	2018-01-28 13:38:02	email	retargeting-campaign	245
10177	2018-01-24 11:58:33	nytimes	getting-to-know-cool-tshirts	232
10503	2018-01-08 01:39:21	buzzfeed	ten-crazy-cool-tshirts-facts	190
10677	2018-01-18 05:07:47	medium	interview-with-cool-tshirts-founder	184
10254	2018-01-25 09:12:18	google	paid-search	178
10925	2018-01-20 11:55:30	google	cool-tshirts-search	60

2.3 How many visitors make a purchase?

Query Results
COUNT (DISTINCT user_id)
361

```
-- You can put your query here
SELECT COUNT (DISTINCT user_id)
FROM page_visits
WHERE page_name = '4 - purchase';
```

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

```
-- You can put your query here
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)
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;
```

Query Results

user_id	last_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
10162	2018-02-01 04:26:10	email	weekly-newsletter	115
10069	2018-01-04 08:13:01	facebook	retargetting-ad	113
10030	2018-01-28 13:38:02	email	retargetting-campaign	54
10354	2018-01-22 02:00:29	google	paid-search	52
22224	2018-01-26 06:30:52	buzzfeed	ten-crazy-cool-tshirts-facts	9
26835	2018-01-07 02:12:04	nytimes	getting-to-know-cool-tshirts	9
14770	2018-01-15 22:54:18	medium	interview-with-cool-tshirts-founder	7
48931	2018-01-17 10:39:53	google	cool-tshirts-search	2

2.5 CoolTShirts can re-invest in 5 campaigns. Given your findings in the project, which should they pick and why?

The first 5 campaigns in the `COUNT(utm_campaign)` column should be chosen, because they are the ones with the most last touches, that is, the most clicks on the purchase page.

Query Results				
user_id	last_touch_at	utm_source	utm_campaign	COUNT(utm_campaign)
10162	2018-02-01 04:26:10	email	weekly-newsletter	115
10069	2018-01-04 08:13:01	facebook	retargetting-ad	113
10030	2018-01-28 13:38:02	email	retargetting-campaign	54
10354	2018-01-22 02:00:29	google	paid-search	52
22224	2018-01-26 06:30:52	buzzfeed	ten-crazy-cool-tshirts-facts	9
26835	2018-01-07 02:12:04	nytimes	getting-to-know-cool-tshirts	9
14770	2018-01-15 22:54:18	medium	interview-with-cool-tshirts-founder	7
48931	2018-01-17 10:39:53	google	cool-tshirts-search	2