code|cademy

Capstone: Attribution

Learn SQL from Scratch
Ugonna Nwaokwu
6/9/2018

Example Table of Contents

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

1. Get familiar with CoolTShirts

1.1 – Get familiar with CoolTShirts

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

CoolTShirts has 8 campaigns and 6 sources.

Please see which source is used for each campaign on screenshot below

```
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;
```

utm_campaign	utm_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

1.2

What pages are on the CoolTShirts website?

There are four pages o the CoolTshirts website

```
Select distinct page_name
From page_visits;
```

Query Results

page_name
1 - landing_page
2 - shopping_cart
3 - checkout
4 - purchase

column1	column2	column3

2. What is the user journey?

2.1 – What is the user journey?

How many first touches is each campaign responsible for?

Please see first touches each campaign is responsible for on screenshot below

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

Query Results

Database Schema

```sql
WITH first_touch AS (SELECT user_id,
       MIN(timestamp) as first_touch_at
    FROM page_visits
    GROUP BY user_id),
ft_attr AS (
  SELECT ft.user_id,
       ft.first_touch_at,
       pv.utm_source,
       pv.utm_campaign
  FROM first_touch ft
  JOIN page_visits pv
    ON ft.user_id = pv.user_id
    AND ft.first_touch_at = pv.timestamp
)
SELECT ft_attr.utm_source,
       ft_attr.utm_campaign,
       COUNT(*)
FROM ft_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

2.2

How many last touches is each campaign responsible for?

Please see last touches each campaign is responsible for on screenshot below

Query Results

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

```
WITH last_touch AS (SELECT user_id,
        MAX(timestamp) as last_touch_at
    FROM page_visits
    GROUP BY user_id),
lt_attr AS (
  SELECT lt.user_id,
        lt.last_touch_at,
        pv.utm_source,
        pv.utm_campaign
  FROM last_touch lt
  JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
)
SELECT lt_attr.utm_source,
        lt_attr.utm_campaign,
        COUNT(*)
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

2.3

How many visitors make a purchase?

361 visitors made a purchase

```
Select count (distinct user_id)
From page_visits
Where page_name = '4 - purchase';
```

column1	column2	column3

2.4

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

Each campaign is responsible for below last touches on the purchase page

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

Query Results

```
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),
lt_attr AS (
  SELECT lt.user_id,
         lt.last_touch_at,
         pv.utm_source,
         pv.utm_campaign
  FROM last_touch lt
  JOIN page_visits pv
    ON lt.user_id = pv.user_id
    AND lt.last_touch_at = pv.timestamp
)
SELECT lt_attr.utm_source,
       lt_attr.utm_campaign,
       COUNT(*)
FROM lt_attr
GROUP BY 1, 2
ORDER BY 3 DESC;
```

3. Optimize the campaign budget

3.1 – Optimize your campaign budget

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

CoolTShirts should re-invest in the below 5 campaigns because they have the highest number of campaigns that led to purchase

lt_attr.utm_campaign	COUNT(*)
weekly-newsletter	115
retargetting-ad	113
retargetting-campaign	54
paid-search	52
ten-crazy-cool-tshirts-facts	9